Policy Playbook on
Automated Support for Decision-making

2021 edition

How to ensure responsible use of automation, advanced analytics and artificial intelligence in service of administrative decision-making

February 2021   *Strategic Policy and Planning Branch*   **A-2021-28399-000001**

PREAMBLE

Why do we need new policy guidance in this domain?

The premise for automating aspects of administrative decision-making at IRCC is familiar: application volumes are rising in temporary resident, permanent resident and citizenship lines of business; operational networks are continuously required to expand their capacity; and clients are increasingly expecting quick and easy digital interaction. IRCC is undergoing a profound transformation, and new uses of technology are seen as one viable avenue to keep pace and still maintain quality.

Big data and artificial intelligence (AI) are exploding, and IRCC is beginning to recognize the potential they bring for new business insights and efficiency gains. But this is a new area, and conventional methodologies, guardrails and governance models are often incompatible with how algorithmic approaches really work. With algorithms and automation growing in importance, the Department would benefit from a considered approach to resolving novel questions about privacy, data governance, security, transparency, procedural fairness, and human-machine interaction.

The Government of Canada has introduced new government-wide policy around responsible and ethical use of automated decision systems, which came into force on April 1, 2020. The Treasury Board Secretariat (TBS) *Directive on Automated Decision-Making*, which supports the *Policy on Service and Digital*, outlines high-level requirements for departments using systems, tools or statistical models to recommend or make administrative decisions about clients. The Directive is be accompanied by a mandatory Algorithmic Impact Assessment tool, which must be initiated at the design phase and finalized before any new system moves from a test environment to live implementation. These overarching policy instruments will help IRCC to navigate uncharted waters.

To be successful, IRCC also needs clear and readily available guidance that takes into account its unique position as a service provider for both Canadians and foreign nationals, as well as its specific operating context, systems and programs. IRCC's internal policy starts with fundamental questions like: *Is the use of automation to support decision-making a good idea in my case? Do we have what we need to be successful? Do the benefits outweigh the risks? How do we do this securely?* It then grounds key technology ethics concepts in our day-to-day work, and directly answer the questions we are asking ourselves, such as: *Could we ever use algorithms to automate negative decisions on client applications? When, if ever, are "black box" algorithms appropriate? Who is accountable for individual decisions made by a machine? How do we meaningfully explain automated decisions to clients? How can we give IRCC's officers the benefit of insights gleaned from analytics without fettering their decision-making?*

IRCC needs to find the right answers to these questions to maintain public confidence, which is critical to successful immigration, refugee, citizenship, settlement and passport programs. Codifying or "hard-wiring" business rules, whether based on advanced analytics or on common sense heuristics long relied upon by individual officers, makes them subject to more scrutiny. The application of business rules at scale elevates the impact if something were to go wrong, requiring additional safeguards within the Department.

On top of managing anticipated risks, specific policy guidance on automated support for decision-making can help IRCC realize the full benefits of new approaches. Principles are intended to enable IRCC staff to consider not only what actions the Department should avoid because they are too risky, but also to what activities we should pursue because they can enable us to deliver *better* immigration programs and services.

What is the scope of activities this playbook targets?

This policy targets **automated systems that support, in whole or in part, administrative decisions.** This includes systems that: classify cases according to the level of scrutiny they require; flag cases for human review or investigation; provide recommendations about whether applications should be approved; or render complete decisions. The business rules applied by automated systems that support decision-making could be derived from sophisticated data analytics, or from interviews with experienced officers – what matters is that these systems take up a new role in IRCC's decision-making model. Analytics-based rules are a particular focus within this policy, especially if machine learning is involved.

This policy does not cover all types of automation. Routine process automation, such as the programming of computers to accomplish repetitive clerical tasks formerly done by humans, is out of scope. So too are systems that merely help to divide our collective caseload between networks, offices or officers for the purpose of efficiency. If the automation does not play a role in helping IRCC to reach an administrative decision that affects a client, it is not of primary concern in this policy. However, staff are encouraged to carefully consider whether automation that seems removed from final decisions may inadvertently contribute to an approval or a refusal.

What shape does the playbook take?

Full spectrum support for ethical conduct means having both a code of personal ethics and a set of institutional ethics. In terms of personal ethics, nothing new is needed. The existing <u>Values and Ethics Code for the Public Sector</u> is domain-agnostic (i.e. it already applies equally to accountants doing internal audits as it does to data scientists developing machine learning systems). The code's five core values of respect for democracy, respect for people, integrity, stewardship and excellence provide a sound and enduring basis for IRCC's innovators to make decisions about how to comport themselves ethically and professionally.

With respect to institutional ethics, a set of guiding principles is proposed in Part I. **As a general rule, principles do not identify specific courses of action, but instead serve as a basis for decision-making when people find themselves in new territory, confronting novel problems**. By disseminating the guiding principles in Part I, IRCC hopes to see its data scientists, program designers and policy developers inspired, trained and empowered to prioritize ethical considerations in the development of automated support for decision-making systems. In particular, they provide guidance to the Department as a whole and its senior managers, who are ultimately accountable for the results of automated systems and for creating a workplace that fosters responsible use of technology. Reflecting upon and openly discussing the wider impacts of the Department's innovative work can only help to ensure that it consistently supports the public good.

The principles presented in this policy are designed as a companion to the more concrete guide contained in Part II. **The Automator's Handbook guides IRCC staff through the primary questions that need to be considered at various stages of a project**: (1) when exploring automated support for decision-making as a possible solution to the problem at hand, (2) when designing a new system, (3) when preparing to launch, and (4) once an automated system is up and running. The handbook is geared toward individuals involved in developing and implementing automated systems that support decision-making. This may include data scientists, program and policy specialists, privacy experts, IT, and operational staff.

The Playbook covers additional chapters including **an overview of legal considerations and practical tips to align with administrative, human rights and privacy law (Part III)** and **baseline privacy requirements (Part IV).** The policy concludes with **a checklist to support compliance with the *Directive on Automated Decision-Making* (Part V)**, and a **glossary of key terms**.

Questions or comments about the Playbook? Contact IRCC's Digital Policy team in the Strategic Policy and Planning Branch: <u>IRCC.Digital-Numerique@cic.gc.ca</u>.

A new policy on automated support for decision-making is an opportunity to ensure that the Department's thinking keeps pace with the speed of technological change, and that our people and practices continue to deliver a suite of programs equal to the expectations of Canadians and the world. The foundational principles and standards that have guided IRCC are not changing. Our commitments, mission and duties remain consistent. But while it may seem that the tools we use in the digital era merely recreate the paper and in-person practices of the past, they do not. Adoption of these technologies has and will change what we do and how we do it, the skills that we need, our relationship with partners and clients, and the culture of our workplace.

Note: This policy should be seen as a living document that needs to be regularly updated over time to ensure continued relevance as the technology, and our knowledge thereof, evolves.

Immigration, Refugees
and Citizenship Canada

Immigration, Réfugiés
et Citoyenneté Canada

Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

PART I – A SET OF GUIDING PRINCIPLES

How should the Department as a whole proceed when faced with difficult new choices about specific applications of emerging technology? The following set of principles outlines IRCC's overarching goals and covers a range of topics, like the right tools for the right circumstances, responsible design, transparency and explainability, and recourse.

*Click the arrows on the left for an expanded treatment of each principle.

Who needs to read this?

- *IRCC's senior managers;*
- *IRCC practitioners working in this domain; and*
- *Anyone interested in learning about the broad parameters that guide IRCC's use of AI and automation.*

1. ***The use of AI and automation should deliver a clear public benefit. IRCC should use these tools wherever it can do so responsibly, effectively and efficiently – in that order.***

1.1 **IRCC's use of automated support for decision-making should build trust in Canada's immigration system**. Maintaining the confidence of Canadians is critical to the Department's success, and IRCC should actively address potential anxiety associated with Big Data and AI. The Department should take measured steps to build legitimacy and public trust around its use of automation, analytics and AI.

1.2 There is an opportunity cost to <u>not</u> pursuing the use of automation, advanced analytics and AI. IRCC should not be so cautious as to miss out on the potential of data-driven technologies. As in any era, **making use of the best tools available is part of effective and efficient use of public money.** In some cases, algorithmic systems are the only efficient way to process the massive amount of information needed for the operation of modern government services.

1.3 However, **IRCC can only employ these technologies to the extent that their use remains consistent with Canadian norms and standards** regarding equality and non-discrimination, procedural fairness, transparency, privacy, security, and accountability. The responsible use of technology must respect requirements of administrative law, and of overarching frameworks like the Charter of Rights and Freedoms and the *Privacy Act.*

1.4 **IRCC should not pursue automating decisions (or aspects of decisions) at the expense of program integrity.** Risk management is at the heart of IRCC's work. Automation may change risks or introduce new ones, but innovation should seek to boost program integrity.

1.5 **IRCC should make every effort to obtain reliable data** in order to make a direct link between clients' outcomes and the information they submit at the application stage. At the same time, **the Department needs to carefully consider the strength of the inferences it draws from new data sets**. Differentiating correlation and causation is a perennial challenge, and human intentions are hard to predict reliably.

1.6 When identifying potential areas to test the use of AI and automation, IRCC should consider the following factors:

— **Need:** Canada has seen year-over-year growth in the volumes of applications received across several lines of business, and expects this trend to continue for the foreseeable future. While growth is a primary driver for the adoption of AI and automation in IRCC's decision-making processes, enhanced program integrity management and rising client expectations are equally important. Focusing on need from the onset will ensure that IRCC is responding to a clearly identified problem and avoiding the introduction of new technologies for the sake of 'going digital';

— **Data:** the larger the traveller movement, the larger the corpus of cases to learn from, and where usage of online applications is widespread, data is more structured. The availability of large volumes of highly-structured data increases the quality of data, which can reduce unintended bias, and increase the integrity of data used to train algorithms and support decision-making. When exploring the use of machine learning to support temporary resident visa (TRV) processing, IRCC looked for caseloads where data integrity was high;

— **Stakes:** the gravity and finality of the decision, including its impact on clients, may vary by line of business. Temporary residence applications are generally considered lower stakes than permanent residence or citizenship applications, and certainly lower stakes than humanitarian and compassionate cases or pre-removal risk assessment; and

— **Complexity:** new technologies are more suitable for applications that involve routine examination of information, rather than cases that require more nuanced judgment calls.

See TBS Directive, section 4 (Objectives and expected results).

2. *Humans (not computer systems) are accountable for decision-making, even when decisions are carried out by automated systems.*

2.1 **The introduction of automated systems can change the *time and place* of human intervention in the decision-making process.** For example, humans may take on new roles, like setting business rules for an automated triage system to carry out. **However, automated systems should not displace the central role of human judgment in decision-making**. There is an essential role for humans in deciding which types of systems to use, which cases to apply them to, and which values to encode. In cases where algorithms play a role in administrative decisions that significantly affect people's lives, it is reasonable for applicants and Canadians to expect that a real person has exercised human judgment in developing the process, and that a person is ultimately accountable for its final results.

2.2 **Maintaining human intervention (often referred to as 'human-in-the-loop') is important for public confidence. Models that incorporate automation should maintain a role for people in the decision-making process.** Human-in-the-loop systems represent a form of transparency and accountability that is more familiar to the public than automated processes. This means that a human decision-maker must be identified as the officer of record for all decisions made or supported by an automated decision system.

A-2021-28399-000006

2.3   Accountabilities cannot be passed to machines when they take over tasks previously done by humans. **IRCC must take ownership for the successes and failures of its systems.**

2.4   IRCC should also reflect on the best approach to **devoting time to questions concerning the responsible use of AI** within its planning, advisory and decision-making structures.

See TBS Directive, sections 6.3.9-6.3.10 (Ensuring human intervention).

3.   *Because IRCC's decisions have significant impacts on the lives of clients and Canadians, the Department should prioritize approaches that carry the least risk.*

3.1   Automating final decisions is the highest risk approach to achieving efficiencies. IRCC should first **explore whether efficiencies can be achieved by automating other aspects of the business flow** (e.g. workload distribution, risk-tiering, note-taking, research, communications, etc.)

3.2   When using a predictive model for decision-making, automating approvals is usually considered less impactful than automating refusals. However, it is important to view the situation from the perspective of individual applicants – particularly those applying to a business line where space is limited. **In situations where demand exceeds supply, the stakes for positive decisions are raised**: it may be perceived that an approval (or invitation or acceptance) of one candidate's application comes at the expense of another's.

3.3   Another perspective is that of Canadians, who accrue the economic, social and cultural benefits of temporary and permanent migration, but who also take on the safety and security risks. **From a collective perspective, automated approvals based on prediction can carry significant consequences**. It is understandable for Canadians to be more concerned about mistakenly approving risky applications than about mistakenly refusing *bona fide* candidates. IRCC needs to take a balanced view, considering fairness from both the individual and collective perspectives.

3.4   Because IRCC cannot fully eliminate risks to the security of its systems, it should **focus on building security measures into the design of its AI models and infrastructure to minimize impact in the event of a breach**. The mandatory Security Assessment and Authorization supports innovators in making sound decisions that enhance system integrity, management of sensitive information, and oversight.

See TBS Directive, sections 4 (Objectives and expected results); 6.1 (Algorithmic impact assessment), and 6.3.7 (Security).

4.   *"Black box" algorithms can be useful, but cannot be the sole determinant of final decisions on client applications.*

4.1   **True "black box" AI tools are of limited utility for administrative decision-making** because the Department should not make decisions it cannot meaningfully explain. "Black box" tools, such as a facial recognition application, can be used in a supporting role – but even then,

Immigration, Refugees Immigration, Réfugiés
and Citizenship Canada et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

Policy Playbook on Automated Support for Decision-making – DRAFT

confidence/reliability levels should be sufficiently high; and where they are not, humans must have an opportunity to review the system's results before final decisions are made.

4.2 **Black box algorithms should be permitted for the purposes of business optimization**, such as determining where to locate offices or how to best distribute resources among them.

See TBS Directive, section 6.2.3 (Providing explanations after decisions).

5. *IRCC must recognize the limitations of data-driven technologies and take all reasonable steps to minimize unintended bias.*

5.1 Every machine-learning algorithm operates entirely within the world defined by the data used to calibrate it, so **limitations or flaws in data sets will bias outcomes**, sometimes severely. In creating tools designed to make decisions on immigration, citizenship and passport applications, **IRCC must be careful not to replicate and "hard wire" historical biases** under a guise of technological neutrality. Understanding our data and planning for its responsible use is not a purely technical question – interdisciplinary collaboration is required to address fairness meaningfully.

5.2 AI-powered tools can also be used to study existing programs in new ways, to ensure they are achieving their intended results, and to **shine a light on any bias that may have inadvertently crept into long-standing practices** of human decision-makers.

5.3 All analytical tools are limited in their ability to predict and describe outcomes. While overall patterns and trends are useful, they cannot substitute for evidence of personal actions. For this reason, **IRCC should be very careful when applying information about a group to an individual**.

5.4 **Big data should be used to help IRCC look backward as well as forward.** IRCC should look to use new tools to detect, assess and remove any bias or inconsistencies.

5.5 **IRCC should constantly take steps to improve the quality of its data holdings, given that good data is foundational to advanced analytics and AI.** Prioritizing the development of data-driven tools requires prioritizing the breadth, depth and reliability of our data sets.

5.6 Analytics may suggest that an application is low-risk, based on how the application profile aligns with that of successful applicants in the past, and in this case **IRCC may wish to take on the risk of an automated approval. However, the calculus changes in the reverse scenario**: an automated refusal cannot be based solely on similarities between an applicant's profile and profiles of unsuccessful past applicants. **Refusals must be based on *evidence* about the particular application in question**. For example, an automated refusal based on a TRV applicant's passport being in Interpol's Stolen and Lost Travel Documents database would be justifiable because it is based on an objective, verifiable fact about the applicant in question (i.e. they do not meet the requirement for a valid travel document). Conversely, it would not be justifiable to base an automated refusal on the fact that the applicant's country of origin has a high incidence of passport fraud – this could serve as a risk indicator amongst others, but it is not itself a sufficient basis for an outright refusal.

See TBS Directive, sections 6.3.1-6.3.2 (Testing and monitoring) and 6.3.3 (Data quality).

6. *Officers should be informed, not led to conclusions.*

6.1 Automated systems that support decision-making can be used to assist officers in exercising their discretion. In these cases, systems must be designed so that they **do not fetter decision-makers** in the exercise of their capacities and authorities.

6.2 **Officers need appropriate training and context to understand what, precisely, a decision support system is telling them**. For example, labels matter: when a system automatically triages clients' applications into groups, it makes a difference whether these groups are labelled "green/yellow/red" or "straightforward/average/complex." The former may appear to represent instructions to officers (green=approve, red=refuse), the latter a simple indication of how much time and effort an officer will need invest to reach a decision.

6.3 Big data should **strengthen administrative decision-making by providing a richness of relevant information to decision-makers, and by focusing their attention on the determinative factors**. In this way, new tools should help officers make not only faster, but *better* decisions.

6.4 **IRCC should endeavour to provide its decision-makers with the best information and tools we have at our disposal**. Putting new insights at the fingertips of decision-makers and focusing their attention on the most relevant factors will reduce wasted effort, boost consistency and better link action and rationale.

6.5 User-testing during the design and development of a project can enable the Department to **develop a strong grasp on how humans interact with automated systems**. Beyond providing insights into the most effective and reliable ways to present machine-generated outputs to officers, a more in-depth understanding of the dynamics at play can enhance accountability and help prevent unintended negative consequences.

See TBS Directive, section 6.3.5 (Employee training).

7. *Humans and AI play complementary roles. IRCC should strive to sharpen the roles of each.*

7.1 Automation, advanced analytics and AI are force multipliers; they can **assist and augment the capabilities of people**. Using these technologies to allow IRCC staff to focus on more creative, problem-solving, or client-facing work should improve both service delivery and job satisfaction.

7.2 **IRCC should leverage the strengths of computers** – reliably analyzing large volumes of data and considering millions of possibilities without fatigue – **and let humans focus on theirs**: intuition, creativity, empathy, social skills, shaping a larger strategy. **Algorithmic systems should allow people to focus on things that require their expertise and judgment.**

7.3 **Systems that make administrative decisions on their own** – as opposed to assisting a human decision-maker – **are suitable for decisions involving straightforward, factual determinations**, such as whether a client has a criminal record or has visited Canada before. Automation should be

focused first and foremost on routine tasks – where desired outcomes are easily described and where human judgment is not required.

7.4 In contrast, when automated systems are intended to assist a human decision-maker, training should be designed in a way that encourages officers to **apply their own professional judgment and intuition**, as a means of augmenting decision-making[1].

7.5 **Having a human involved in the process not only safeguards against procedural fairness risks, it gets better results.** For straightforward types of assessments (e.g. photo matching), the performance of data-driven algorithms is often superior to that of humans. But when machines and humans team up, the results are better still. One way to boost performance is to have experts apply judgment to the output of mathematical models; another is to flip the sequence and have program experts and decision-makers contribute to the design of the model in the first place.

7.6 **IRCC should regularly review and adjust its automated systems** to ensure they reflect the reality on the ground and are operating in alignment with the policy priorities of the day. Feedback loops are critical to responsible design and quality control. For example, feedback from frontline officers must be able to reach system developers, so that IRCC can constantly reassess whether a system's conclusions are reliable. **It is essential that officers continue to review a portion of applications manually**. This will help to ensure that officer skills do not atrophy and that knowledge about country conditions and fraud trends remains up-to-date.

7.7 **Technical experts (and their managers) are in a position of trust, and have a special responsibility to thoroughly evaluate the business logic and the potential issues that may arise due to automation.** It is imperative that the Department receive objective assessments and frank advice. With the advent of machine learning systems, much care should be taken to identify, explain and mitigate potential risks, especially when risks may only be apparent to those with a technical understanding of those systems. Managers need to create conditions that allow experts to consult peers, reassess risks as systems evolve and report hazards without fear of reprisal.

7.8 IRCC should make every effort to **give its employees the skills, tools and support they need** to pursue innovation. Support includes time to learn, and the use of secure sandboxes or testing environments that allow employees to experiment responsibly.

See TBS Directive, sections 6.3.9-6.3.10 (Ensuring human intervention) and 6.3.5 (Employee training).

8. *IRCC should continually adopt emerging privacy-related best practices in a rapidly evolving field.*

8.1 Data scientists, program designers and IT specialists should be conversant in privacy issues, including **understanding the rights and responsibilities associated with the collection, use, disclosure and retention of personal information**, as well as related **best practices and standards** IRCC should give

---

[1] https://www.nesta.org.uk/blog/human-vs-machine-why-we-need-be-thinking-more-about-how-humans-are-interacting-ai-tools/

A-2021-28399-000010

thought to developing tools and procedures to assist staff in assessing and addressing privacy needs in the age of big data.

8.2 The Department should **be proactive in preparing for a possible privacy breach or investigation into how it manages personal information** in its algorithmic systems. This means following appropriate security and privacy guidelines regarding breaches, and ensuring that training is current. In the event of a complaint or review by the Office of the Privacy Commissioner, IRCC must be prepared to produce copies of all documentation outlining the steps taken to comply with privacy requirements, and steps taken to exceed mere compliance, where appropriate.

See TBS Directive, section 6.3.3 (Data quality).

9. IRCC should subject systems to ongoing oversight, to ensure they are technically sound, consistent with legal and policy authorities, fair, and functioning as intended.

9.1 Inviting **external scrutiny** of the design of our systems is an important aspect of ensuring legitimacy. Where possible, new algorithmic models should be made available to external experts, such as an advisory body, for auditing, testing and review.

9.2 **Quality control** processes should be established early on and conducted regularly. If testing is done internally, the methods and assumptions used, along with the results, should be openly documented and made available to the advisory body. Subjecting automated decision systems to greater control mechanisms immediately prior to deployment and on an ongoing basis post-deployment while providing a lighter touch during exploration will ensure adequate oversight without stifling innovation.

See TBS Directive, sections 6.3.2 (Testing and monitoring) and 6.3.4 (Peer review).

10. IRCC must always be able to provide a meaningful explanation of decisions made on client applications.

10.1 Regardless of the degree to which IRCC relies on algorithmic models to support decision-making, applicants will always have the right to understand the basis of decisions made on their applications. Explanations play a key role in allowing applicants to pursue recourse options, and to government accountability and public trust more generally. Explaining how algorithmic systems work can be technically challenging, but **explanations of automated decisions should not hinge on the general public's understanding of how algorithmic systems function**. If an explanation provides little meaningful information, it offers little value.

10.2 One way to make explanations understandable is to limit a system's performance to those functions that can be easily explained. This may, however, mean squandering the potential of AI systems that can make sense of complex data in ways that humans cannot. **Rather than imposing sharp limits on what new tools can** *do*, **IRCC should endeavour to find new and better ways to** *explain* **them.**

11

Immigration, Refugees   Immigration, Réfugiés
and Citizenship Canada   et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

10.3 Given a prohibition on black box systems for administrative decision-making in full, the use of automation to support decision-making should make decisions **more explainable, not less**. For example, audit trails capturing every factor and step in the decision-making process can be used, where needed, to systematically generate and record reasons for acceptance or refusal.

10.4 An **audit trail** is critical to respecting an applicant's right to understand the basis of the decision on their application, and may be needed to satisfy the courts in the context of a legal challenge. Any new tool that lacks the capacity to **automatically generate a comprehensive audit trail of the administrative decision-making path** is not yet ready for implementation.

See TBS Directive, section 6.2.3 (Providing explanations after decisions).

11. IRCC must be transparent about its use of AI. It must provide meaningful access to the system while protecting the safety and security of Canadians.

11.1 **IRCC needs to find ways to report, in a meaningful way, on the workings of whole systems.** At a minimum:

— Clients need to understand how their applications will be processed and their personal information used;

— Canadians need to understand how IRCC spends public money to facilitate the entry of family, friends and legitimate visitors, while at the same time protecting against threats to health, safety and security; and

— Stakeholders, media and technical experts need a basis for an informed critique of Canada's approach. This means that IRCC should proactively share information about how its models work and, more importantly, how the Department is – and is not – using new technologies to support decision-making.

11.2 The Department has an interest in not disclosing too much information about its decision systems for program integrity reasons. Divulging too much information can allow individuals to manipulate decision-making, or "game the system." Explanations of how systems work, and of the individual decisions they help the Department to make, should be **detailed enough to inform a client generally about how decision will be made, but not so comprehensive as to jeopardize program integrity and, by extension, the safety and security of Canadians.**

11.3 **IRCC should endeavour to design systems, to the greatest extent possible, around factors/data points that are difficult to falsify or misrepresent.** For example, bank records are much more difficult to fabricate than a reason for travel, which is simply stated by the applicant. We should rely on objective facts as much as possible in order to mitigate potential fraud risk. The weight assigned to different data elements in IRCC's algorithms should account for the degree to which they may have been targeted for fraud in the past, or are likely to be targeted going forward. For example, potential security risk factors would be highly weighted.

See TBS Directive, sections 6.2.1-6.2.2 (Providing notice before decisions) and 6.2.6 (Release of source code).

12. IRCC's use of automated systems must not diminish a person's ability to pursue recourse.

12.1 When a client challenges a decision that was made with the assistance of an automated system, and IRCC has the authority and agrees to review the decision, **a human officer will conduct the review**. Similarly, under the Privacy Act, individuals have a right to recourse – via a complaint to the Office of the Privacy Commissioner – if they feel their personal information was mishandled.

12.2 However, **IRCC should not proactively offer clients the choice to have a human officer review and decide on their case in lieu of the automated system at the beginning of the application process**. IRCC's objective is to implement responsible, effective and efficient systems. Allowing clients to opt out of them would compromise the Department's ability to ensure consistent and fast processing, while also erroneously implying that its automated systems provide inferior service.

See TBS Directive, section 6.4 (Recourse).

Immigration, Refugees   Immigration, Réfugiés
and Citizenship Canada   et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

PART II – THE AUTOMATOR'S HANDBOOK

This practical handbook is intended to guide IRCC staff through key questions that should be considered at various stages of AI and automation projects – from early exploration to ongoing monitoring once a system is up and running.

Records of key decisions about the use of AI and automation is important to demonstrating that you have carefully considered a host of factors and identified adequate mitigation strategies along the way. Keeping notes of observations, analysis and links to key documents throughout the design, development and deployment phases can be helpful to support the answers you will need to provide when you complete an Algorithmic Impact Assessment (AIA), and may also need to be produced in the event of litigation. More information about the AIA tool and requirements is provided in section B, *Designing your system.*

Who needs to read this?

- *Individuals and teams who are considering the application of AI and automation to support administrative decision-making. This may include data scientists, policy and program leads, privacy and data governance professionals, IT specialists, and operational guidance experts – most likely at the Analyst, Officer, Advisor, and Manager levels.*

A. Exploring automated support for decision-making as a potential solution

This section will help to work through the key questions that need to be considered at the outset of a project. The answers to these questions should help you determine whether or not an automated decision system is a suitable solution to the problem you are trying to tackle, whether or not you have what you need to be successful, and if the benefits outweigh the risks.

General suitability

Have you clearly defined the problem at hand?

Spending time focusing on the problem before jumping to possible solutions will help avoid 'technology theatre' and make sure that the solution you ultimately choose responds to a clear need. It can also save time and help you identify relevant metrics to measure success.

Do all stakeholders have the same understanding of the problem? Is it an issue that can be solved through AI and automation?

> *Add observations, analysis and links to key documents here.*

Have you exhausted straightforward options before moving into the more complex and contentious world of automated support for decision-making?

In some cases, automation of simple and repetitive tasks (e.g. through the use of robotic process

automation) or new approaches to workload distribution may be able to achieve similar efficiencies. Considering multiple options upfront.

> *Add observations, analysis and links to key documents here.*

Does the nature of the decision or specific step in the decision-making process lend itself to automated decision-making?

When determining whether a task or step in a decision-making process could be shifted from humans to machines, you should ask yourself, "Is this a situation in which reasonable minds may differ?" If yes, automation is ill-advised. Conversely, if analysis of past decisions has shown that virtually any officer would reach the same conclusion given the facts at hand, automation should be pursued.

> *Add observations, analysis and links to key documents here.*

Is this a project that requires special attention? For example:

- It is within an area of intense public scrutiny (e.g. because of privacy concerns) and/or frequent litigation
- Clients in this line of business are particularly vulnerable
- The stakes of the decisions are very high, e.g. you are thinking of automating refusals, or outcomes will have a significant impact on peoples' lives
- It will have a major impact on IRCC staff, either in terms of their numbers or their roles
- You are thinking of collecting/using non-traditional and potentially sensitive data, such as data associated with clients' social media accounts

If one or more of these points applies, your project will require more consultation and oversight at all stages.

> *Add observations, analysis and links to key documents here.*

Is the technology you are thinking of using likely to raise public concern? For example:

- The specific type of technology is a source of negative media coverage or associated with outcomes that are less effective when it comes to certain minority groups (e.g. facial recognition)
- You are thinking of using a "black box" algorithm
  - *How do you know if you have a "black box"? The system's process of analysis or reasoning is unknowable, or its behaviour is difficult to interrupt or explain*

If so, you will need to consider how you will ensure that the introduction of these tools does not negatively impact confidence in IRCC's ability to make decisions that are fair and reliable.

> *Add observations, analysis and links to key documents here.*

Preliminary diagnostics and impact assessments

What impacts, positive and negative, will the proposed automation have on clients?

For example, the project is expected to:
–   Lead to faster processing
–   Require clients to entrust IRCC with more personal information, or otherwise affect their privacy
–   Improve or degrade the quality of information provided to clients with respect to decisions rendered
–   Result in some cases or applications receiving little or no scrutiny by a human decision-maker

> *Add observations, analysis and links to key documents here.*

Will your automated decision system impact diverse groups?

Gender-Based Analysis Plus (GBA+) is a comprehensive analytical tool that helps fully assess the potential impacts of new government policies and initiatives on diverse groups of people. The "plus" acknowledges multiple, intersecting factors such as: sex and gender, indigenous status, income, employment status, family status, education, immigration status and country of birth, age, visible minority status, language, region of residence, disability status.

Applying a GBA+ lens at the very start of your project will help you assess risks that may result in differential impacts for diverse groups. At this stage, you should focus on gathering disaggregated data about clients, analyzing this data for quality and historical bias, and checking your assumptions. AI can also be useful to establish relevant patterns of intersectionality, which may go unnoticed by humans. GBA+ analysis will also come in to play at later stages of your project!

This is the time to engage IRCC's GBA+ Unit, located in the Strategic Policy and Planning Branch, as early as possible, for advice and available training.

> *Add observations, analysis and links to key documents here.*

What impacts, positive and negative, will the proposed automation have on the Department?

For example:

–   Reducing existing backlog of cases
–   Saving money (in the short, medium or long term)
–   Improving consistency and overall quality of decisions
–   Improving program integrity
–   Performing tasks that humans could not accomplish in a reasonable amount of time, allowing officers to focus on more complex tasks requiring human judgement
–   Increasing the risk of litigation or privacy breaches
–   Increasing (or decreasing) the robustness of data being collected and used

> – Changing the role of officers and decision-makers, including their core competencies, job descriptions or classifications, or changing the overall size or make-up of the staff required to deliver a program

> *Add observations, analysis and links to key documents here.*

Have you considered how frontline staff might react?

You may wish to consult with them, and with Human Resources. Even proposals that do not entail substantive changes to officer roles can inadvertently provoke uncertainty and trepidation. Efforts to communicate project goals up-front could reassure affected staff, saving time and effort later.

Consider how you will consult with frontline staff to get their feedback on a proposed automation tool. This would provide an opportunity to convey the objectives of the proposed automation tool and could help ensure concerns can be addressed as early as possible. The section on adopting a User-Centered Approach offers more detailed guidance.

> *Add observations, analysis and links to key documents here.*

Have you identified potential legal implications and risks?

In the initial stages of your project, you will need to establish a diagnostic of legal issues and flag potential legal implications for further review. This should include all areas of public law (i.e. administrative law, human rights law and privacy law). See the guidance below on *Legal Considerations and Tips* for more information on how to assess and mitigate legal risk throughout the AI lifecycle.

IRCC's Legal Services Unit has a team of counsel specialized in questions related to the use of technologies like AI and automation. You must contact Legal Services early in the process and on an ongoing basis to seek advice pertaining to your project.

> *Add observations, analysis and links to key documents here.*

Will you require new policy or legal authorities or regulations?

The Department has broad legal authority to use electronic tools to deliver on its mandate related to immigration and passport decision-making (see section 4.1 of IRPA and amendments to section 2.2 of the Canadian Passport Order), but you could still require new authorities if your project would entail automation in other lines of business or substantive changes to the following: program requirements (e.g. eligibility criteria); program outcomes (e.g. which clients are approved); the collection, use or disclosure of personal information; or the role of officers in the decision-making process. Even if you have the legal authorities to introduce the proposed automation, regulations may be needed to provide guardrails or to specify exactly how and when automation will be used.

A-2021-28399-000017

Immigration, Refugees
and Citizenship Canada

Immigration, Réfugiés
et Citoyenneté Canada

Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

You may need to work with legal and policy leads to seek new authorities or to introduce regulations through a Memorandum to Cabinet or other means.

> *Add observations, analysis and links to key documents here.*

Does IRCC have the breadth and quality of data required to make the proposal workable?

Establishing a diagnostic related to data management will not only help you identify and mitigate legal risk, it will also help you assess the feasibility of your project. In doing so, you should consider data management as it relates to collection, use, disclose, retention, and security of the data, including personal information (which may require a privacy assessment). You should also consider whether you will need to develop a data strategy for the project.

Some of the questions to consider are:

— Is there enough high-quality, structured data to support reliable inferences? What are the limitations of the data? Is it accurate, up-to-date and truly representative, and does it contain historical biases that could be perpetuated? You will need to get an assessment of the quality of your data set, with leadership from the departmental data governance function and collaboration between business and IT. This assessment should be approved by a data governance management table, such as the Data Executive Steering Committee. *CDO Branch to confirm if this is still valid/up-to-date.*
— How many of your data points are proxies for the actual criteria or outcomes you are targeting? All of them? You will need tried and tested indicators that applicants comply with program objectives/requirements.
— Can you use the data? Does the model make inferences about personal information not provided by the applicant? Does it invade or pose significant risks to individuals' right to privacy? On what basis would you use the data underpinning your analytics in a real world application? Do you only need access to information that IRCC already collects (cross program uses of information previously collected for one program should be reviewed), or are you seeking to pull in external data, such as that held by federal, provincial or international partners?
— How will the data be collected and stored, in relation to other IRCC data systems and data assets?

You will need to get advice from the CDO Branch, the ATIP Division and possibly Legal Services about whether your proposed collection, use or disclosure of the data is consistent with the privacy framework governing the Department, or whether new authorities should be sought. The GBA+ Unit can also provide advice around the collection and display of sex or gender information.

> *Add observations, analysis and links to key documents here.*

Have you assessed the privacy needs related to your project?

Generally speaking, you will need to consider what personal information may be used within your project and how it could be used with the lowest privacy interest. You will also need to evaluate data flows to ensure collection, use and disclosures of personal information meets legal and policy

requirements. Familiarizing yourself with 'Privacy by Design' principles will help you think about the possible privacy considerations and best practices that could help make your proposal robust.

IRCC's Integrated Corporate Business Branch has developed new internal tools and procedures to enable IRCC staff to adopt a "privacy sensitive" approach to projects, including those that involve AI, advanced analytics and automation.

Completing a **Privacy Needs Assessment** (internal link) is the first step to determine what measures you will need to take to adequately mitigate and manage privacy risks related to your project. Questions focus on legal authorities for collection, specific types of personal information that will be collected or created, who will store and have access to this personal information, and how the data will be used (e.g. cloud computing, chatbots, analytics, AI, web scraping, data matching). Once you have completed the form, an ATIP advisor will review it and may recommend:

- The development of a new Privacy Impact Assessment or PIA (internal link);
- An amendment or update to one or more existing PIAs; or
- Other privacy work, as appropriate.

> *Add observations, analysis and links to key documents here.*

Training

Have you taken advantage of available training related to data-driven technologies and their effects?

This may range from courses on digital government – such as those offered by the Canada School of Public Service's Digital Academy – to training on privacy and data literacy. Training on GBA+ will also help you unpack questions around bias and identify actions to avoid negative unintended consequences on diverse groups.

While not all members of a team working on an automated decision system will need the same level of expertise in all areas, the TBS *Directive on Automated Decision-Making* requires that adequate training be provided to employees in the design, function, and implementation stages to ensure that they are able to review, explain and oversee the operations of an automated decision system.

> *Add observations, analysis and links to key documents here.*

Partner and stakeholder engagement

Have you thought about which internal subject matter experts needs to be engaged in the project?

Ensuring the responsible use of automation in decision-making requires vast expertise in numerous areas. Bringing together a multi-disciplinary group will help you tackle the policy, the legal, the data, the IT, and the operational questions all once. It will also help you identify relevant strategies to mitigate risk (e.g. bias in your automated system), create greater awareness about your project throughout the Department, and facilitate alignment with other related initiatives.

You will likely need to involve subject matter experts from the following teams:

- **Advanced Analytics Solutions Centre (A²SC), Operational Planning and Performance Branch** – in most cases, A²SC's team of in-house experts in applied data science will be responsible for designing and testing automated models that support decision-making.
- **Legal Services** – tech counsel can provide advice on program requirements, privacy, Charter, administrative law, intellectual property and litigation risks. Under the TBS Directive, you are required to consult with legal services from the concept stage of a project to ensure that the use of the automated decision system is compliant with applicable legal requirements.
- **ATIP Division, Integrated Corporate Business Branch** – the Privacy, Policy and Governance Unit can provide advice on privacy considerations and requirements, including how to complete a Privacy Needs Assessment, whether you need to undertake or revise an existing Privacy Impact Assessment (PIA), privacy notice statement or Personal Information Bank (PIB). The team can also work with you to build privacy into individual projects.
- **Digital Policy team, Strategic Policy and Planning Branch** – this team can provide advice on policy and legal authorities, IRCC's and TBS' policies on the responsible use of AI and automation, connections with other departmental initiatives, etc.
- **GBA+ team, Strategic Policy and Planning Branch**
- **A program policy branch** – depending on the use case, you may need to work with program policy areas to ensure alignment between your initiative and future policy directions, and seek new authorities (e.g. Immigration Branch, Citizenship Branch, Settlement and Integration Policy Branch).
- **Chief Data Office Branch** – as the departmental lead for 'data quality by design,' this branch can help assess data quality and its suitability for your project, and provide advice on questions related to data governance (e.g. data quality, data management, data lifecycle, and business glossaries). This team should be involved in the design, development and implementation of automated systems.
- **Immigration Program Guidance Branch, Integrity Risk Management Branch, and processing networks** – these teams can provide first-hand knowledge of decision-making processes, the role of officers, related guidance, and operational impacts related to your project.
- **Digital Strategy Branch** – this team can help assess how your initiative fits within IRCC's digital IT strategy and business capabilities.
- **Communications Branch** – this team can support you in determining whether there is a need to do any public communications regarding your project.
- **Client Experience Branch** – this team can support the development of end-to-end human-centered design and usability testing related to your project.
- **Human Resources Branch** – this team can assess whether a new automated system will require modifications to generic work descriptions, statements of merit criteria or position classification requirements. HRB can also provide advice on when and how to engage unions.
- **IT Operations, Transformation and Digital Solutions Sector (TDSS)** – this branch is responsible for system design and architecture, and provides advice to meet the IT security needs of the Department, ensure compliance to applicable laws, policies and standards. You will need to engage the IT Security Unit during the design phase to initiate a Security Assessment and Authorization process, which is mandatory under the Policy on Government Security.

Given the scope and complexity that may accompany automation proposals, you will likely also want to think about how you will engage your management team and other senior officials throughout the development of your initiative. This will help ensure senior management has all the information they need to make decisions at the right times throughout the process.

> *Add observations, analysis and links to key documents here.*

Have you thought about whether you need involvement from other government departments (e.g. CBSA) or provincial/territorial governments? Although this policy targets IRCC activities, your project could have impacts on partners (who may also have their own policies/approaches regarding analytics, AI and automation).

> *Add observations, analysis and links to key documents here.*

Will you engage external stakeholders?

External engagement around IRCC's use of AI and automation in decision-making is vital to building public trust, ensuring that there is not a vast disconnect between how IRCC uses these technologies and what the public views as permissible, and, ultimately, enhancing our capacity to advance responsibly in this domain.

It may not be practical to consult external stakeholders on the development of every algorithm or automated system used by IRCC, but they can be consulted on the potential impacts. It is important to consider the perspectives of those who will be affected by automating aspects of the decision-making process, and particularly when new algorithms/systems are part of an entirely new program or a substantial change in delivery approach.

Traditional IRCC stakeholders are still important, but you will likely need to consider a broader array of stakeholders, such as those working in the field of AI/technology ethics.

It is recommended that you deliberately seek views from a diverse group of stakeholders, and document their perspectives as you would when developing a significant policy or legislative change.

> *Add observations, analysis and links to key documents here.*

Planning for design

Have you considered what resources will need to be budgeted for data analytics experimentation and iterative systems development in your Memorandum to Cabinet and/or Treasury Board Submission (if required)?

Uncovering new business insights through the use of advanced analytics takes time and costs money. Building a new decision support system can be a process of trial and error, and budgeting should provide the flexibility to iterate as needed before achieving a final product. Once built, automated systems require resources for maintenance and periodic renewal.

In the early stages of the project, you will likely need to present the proposed approach to the Business Intake Board (BIB). Led by Projects Branch, this centralized intake process for IM/IT investment proposals (and some non IM/IT projects) assesses the business value and prioritization of new ideas

Immigration, Refugees Immigration, Réfugiés
and Citizenship Canada et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

from a departmental perspective. It also helps ensure that new investments support the strategic direction set out by IRCC's priorities and transformation roadmap.

> *Add observations, analysis and links to key documents here.*

Have you considered how resource requirements will change over time?

It is common to need more human and financial resources at the outset, when a rigorous quality assurance regime is vital to ensuring the system is functioning as intended. It is also important to forecast ongoing costs related to system maintenance, monitoring, and renewal.

> *Add observations, analysis and links to key documents here.*

Will your automated decision system be designed in-house or by an external supplier?

In most cases, the design and development of models will likely be done in-house, by A^2SC. In the event that projects are out-sourced to third parties, you should consult A^2SC for advice and technical expertise. You should also review Government of Canada guidance on using third-party vendors when it comes to AI and automation (e.g. Guideline on Service and Digital)

Public Services and Procurement Canada and TBS have established a list of pre-qualified suppliers interested in providing AI solutions to the Government of Canada. You will also need to carefully consider TBS requirements related to access and propriety of source code.

> *Add observations, analysis and links to key documents here.*

User-centered approach

While human accountability and human-in-the-loop models are familiar concepts in the domain of automated decision-making, the idea of an end-to-end user-centered approach is less so.

And yet, failure to consider the needs and realities of users can increase risks ranging from legal liability (as a result of undue reliance or fettering discretion) to diminished returns on investment (because new systems are not well integrated into existing operating practices), and lower-quality decisions (because users lack an adequate understanding of the system and its outputs).

Have you taken steps to understand users' needs and their current environment?

Prior to undertaking a project that could have a substantial impact on the role of officers in decision-making, you will need to have a good understanding of the general operating environment in question.

— Establish who the primary users are and engage them directly to understand their needs. The users of IRCC's automated systems will generally be officers in our operational networks. Use a combination of open calls for comments and targeted solicitation to gather diverse views.
— Identify existing pressures, pain points and limitations, and factors that might support or impede change. This may include constraints related to time, resources or IT infrastructure.

— Anticipate impacts (both perceived and real) of automation on users, recognizing that all users are unlikely to be affected in the same way.
— Familiarize yourself with the tools and processes that are already in use (or planned) throughout the decision-making process and in targeted offices and/or lines of business. This will help ensure that you can develop compatibility and align your system in a way that makes sense to users.
— Consider your own beliefs and potential biases about the current operating environment, and how they may affect your interaction with users. This should be included as part of your broader GBA+ analysis.

Add observations, analysis and links to key documents here.

Immigration, Refugees Immigration, Réfugiés
and Citizenship Canada et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

Policy Playbook on Automated Support for Decision-making – DRAFT

B. Designing your system

This section will help to answer questions like: *How do we design for fairness? How do we safeguard the impartiality of our decision-makers? How do we meet client expectations for transparency and protection of personal information?*

You will also need to provide some of the answers when you complete the Algorithmic Impact Assessment (AIA) – more information about the AIA is provided below.

Model suitability

Have you modeled more than one option?

It is unlikely that there is only one way to achieve your goal. Exploring, mocking up and testing some alternatives might confirm your initial hypothesis and strengthen your business case – or it might help you to uncover an even better approach.

Remember that the privacy framework applies even in the modeling stage of your activities. The ATIP Division can provide information on specific requirements, and advice to help you adopt a privacy-sensitive approach.

> *Add observations, analysis and links to key documents here.*

Have you developed a detailed process map that will enable policy specialists, privacy experts, legal counsel and others outside of your domain to understand the business process in question, and how, specifically, the introduction of automation will change it?

> *Add observations, analysis and links to key documents here.*

Algorithmic Impact Assessment (AIA)

Have you completed a preliminary AIA?
The AIA is a questionnaire designed to help you assess and mitigate risks associated with deploying an automated decision system. It includes roughly 60 questions about the project, algorithm and data, as well as de-risking and mitigation measures for data quality, procedural fairness and privacy.

Completing a preliminary assessment at the design stage will help you anticipate risks associated with your project, identify mitigation measures and plan to comply with requirements related to:

- Peer review
- Notice
- Human-in-the-loop for decisions
- Explanation requirements
- Testing
- Monitoring
- Training
- Contingency planning

Immigration, Refugees Immigration, Réfugiés
and Citizenship Canada et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

- Approval for the system to operate

Determining the Department's responses to the questions in the AIA will require the perspectives of multiple internal subject matter experts. As such, you should engage policy and program representatives, legal counsel, data governance and data scientists to complete this assessment, and keep records of any supporting documentation or analysis. You may also need to obtain approval from certain branches (e.g. CDO Branch signoff on questions about input data and data quality).

You will need to update the AIA to reflect any substantial changes to your model, and release the final results publicly.

> *Add observations, analysis and links to key documents here.*

User-centered design

Have you engaged officers in designing your system's user interface and thinking about the most effective approach for human involvement?

Getting the most out of automation tools requires thinking about both the most appropriate way to use technology and about the best way to involve human. Focusing specifically on human-machine interaction will help strengthen accountability, safeguard the impartiality of decision-makers, and strike the right balance between the roles of humans and the system. Engaging users directly in the design of these processes (rather than simply informing them after the fact) will build trust and confidence in the automated system. Effective co-design should also draw on rigorous experimentation and an iterative approach – i.e. start small, test different approaches, gather insights, adapt the design, and repeat.

— Experiment with different ways of presenting machine information to officers, using a rigorous approach to measure the impact of various designs. For example, using colour-coded labels to triage applications may prompt different types of responses than presenting information labelled with a percentage or level of complexity. Poorly designed machine outputs and human-machine interaction can mask issues of automation bias and fettering decision-making.
— Give careful consideration to ensuring that the system's user interface and outputs encourage users to exercise their own judgment.
— Work with officers and other multidisciplinary team members to map out the key touch points and tasks requiring human intervention. You should also consider the circumstances in which an officer would be authorized to override a decision made by the system.

— Design and deliver training that gives officers a full understanding of the capabilities and limitations of the system.

> *Add observations, analysis and links to key documents here.*

Fairness and non-discrimination

What steps will you take to minimize unintended bias?

Poorly designed algorithms can introduce bias, but in most cases bias comes from the data that is used for analytics or for the training of machine learning models. Program designers need to be aware of weaknesses in the data at the root of their project. Some techniques for mitigating bias in data include consulting subject matter experts to ensure the data is situated in the appropriate context (e.g. has IRCC changed any business definitions over time?), and subjecting machine learning models to adversarial training that could correct for perceived unfairness. You may also want to consider having external partners or experts review your proposed data sets to help identify any gaps or possible concerns. Rigorous quality assurance testing of results can also detect trends that suggest bias, allowing for immediate correction.

Whatever the specifics of a given project, it is always advisable to assemble a diversity of views and identities within your development team. There are numerous examples of AI projects gone wrong due to a lack of diversity within the brain trust. Overlooking the unique circumstances faced by different groups of people can have damaging impacts on the populations your system is designed to assist. Guarding against unwanted bias means bringing a GBA+ lens to your data analysis and early design work. You should endeavour, to the greatest extent possible, to have team members who bring different worldviews, be they based on sex, gender, age, ethnicity, language, rurality, ability or other identity factors.

Add observations, analysis and links to key documents here.

How will your system be designed to ensure procedural fairness?

In order to preserve (or enhance) procedural fairness, you will need to think carefully about how the addition of automation will change application processing and decision-making. For instance:

- If you are planning a predictive model, consider how it will inform officers without fettering their discretion.
- Ensure that all relevant factors are still being considered.
- If your system will draw on information beyond what an applicant submits, consider how clients will be informed. Clients have a right to know what other ("extrinsic") evidence was used in decision-making, in part so they have the opportunity to correct any erroneous information.
- Establish the reasons that will be provided when a decision is made.

Add observations, analysis and links to key documents here.

Explainability and transparency

How will you ensure that meaningful explanations are provided to clients?

The Government of Canada's *Directive on Automated Decision-Making* requires that meaningful explanations be provided to affected individuals of how and why the decision was made. Specific parameters vary depending on the level of impact associated with the automated system.

As a general rule, explanations of automated decisions should: (1) help clients understand why a particular decision was reached, and (2) provide grounds to contest the decision should the client wish to do so. Program designers should also consider whether it is possible to indicate, at a high level, what would need to change in order for the client to achieve a desired result in the future. This is sometimes called a "counterfactual explanation."

IRCC has yet to pursue automated refusals of client applications. If the Department were to do so, we would need to be mindful of the following:

- Administrative decisions made by, or with the assistance of, an automated system must be at least as explainable as the decisions traditionally made by humans alone. This is why "black box" algorithms, like advanced neural nets, are not recommended as a primary way of determining whether to approve an application.
- Explanations must not jeopardize program integrity. A level of transparency that would include disclosing the inner workings of an AI system is not generally necessary. Instead, explanations are mainly required to answer questions like: What was the basis of a decision? What were the main factors? Would changing a certain factor have changed the decision? Why did two similar-looking cases lead to different decisions?

> *Add observations, analysis and links to key documents here.*

Considering the need for transparency, particularly in the context of legal challenges, are you proposing business rules that are defensible and difficult for individuals to "game"?

- Models and rules need to address program requirements such that decisions are based on these requirements and on relevant, reliable information.
- Design your system, to the greatest extent possible, around factors/data points that are difficult to falsify or misrepresent. Much of the information on client applications is simply stated by applicants, with no way for IRCC to authenticate, but other data points, such as bank records or previously issued visas, are verifiable. Preference for the latter will help mitigate the risk of "gaming by claiming" on the part of applicants who start out with knowledge of how IRCC's systems work.
- If weighting variables in an algorithm, your weighting should account for the degree to which each variable may have been targeted for fraud in the past, or is likely to be targeted in the future.

> *Add observations, analysis and links to key documents here.*

Privacy

How will your system uphold the privacy rights of clients and Canadians?

Consider Privacy by Design principles throughout development to help ensure privacy is built in from the start.

Your approach to using personal information for legitimate ends, without violating the rights of individuals or groups, should include:
— Understanding the provenance of the data
— Limiting the collection of personal information to only what is necessary
— Ensuring its accuracy
— Reducing unnecessary, redundant or marginal data
— Taking precautions to prevent re-identification of anonymized (or pseudo-anonymized) data or unauthorized data collection
— Protecting data from unauthorized access and accidental disclosure
— Ensuring that plans are in place to mitigate damage in the event of a security breach, unauthorized access or accidental disclosure

You should draft transparent data policies and privacy procedures for your project to allow individuals to:

— Understand and manage what data is being collected, how it is being used and disclosed, and how long it will be kept
— Give meaningful consent for automatic data collection, and for any after-the-fact data sharing
— Review, obtain and correct inaccuracies in their personal data

To ensure that privacy is considered at every stage of IRCC's advanced analytics projects, a privacy expert works within advanced analytics teams. Beyond supporting a more seamless process, this means that IRCC's privacy experts are better equipped to provide guidance related to the unique challenges and opportunities when it comes to AI and automation. The Government Advisory Directorate at the Office of the Privacy Commissioner (OPC) can also provide advice on privacy matters related to your project, even if you are not undertaking a Privacy Impact Assessment. To do so, start by contact the ATIP Division, IRCC's central point of contact for dealing with the OPC.

> *Add observations, analysis and links to key documents here.*

Working in the open

Have you considered whether information of the work you are doing should be shared externally?

While the use of AI in decision-making is an emerging and rapidly evolving domain, chances are, other teams are thinking about similar issues and facing similar problems. While data, personal information and coding of a sensitive nature still needs to be protected, there may be other types of information related to your project that could be shared openly and reused by others.

If you have conducted research, prepared reports or developed guidance about AI and automation that could be released without compromising privacy, security or program integrity, consider whether it

A-2021-28399-000028

could be made available through IRCC's website or another open source development platform, like GitHub or the Government's Open Data portal.

> *Add observations, analysis and links to key documents here.*

Will your system use open source software?

The Policy on Service and Digital requires departments to use open standards and open source software by default, with exceptions for circumstances where an open source option is not available or does not meet user needs.

If not using open source, departments are expected to favour platform-agnostic commercial off-the-shelf (COTS) options over proprietary COTS, avoiding technology dependency and allowing for substitutability and interoperability.

If developing a custom-built application, by default any source code written by the government must be released in an open format via Government of Canada websites and services designated by TBS.

See requirements set out in the TBS *Directive on Automated Decision-Making* around access to components and release of source code, and also the guidance provided in the Guideline on Service and Digital.

> *Add observations, analysis and links to key documents here.*

Accountability

Who will be the decision-maker of record?

IRPA Section 4.1 specifies that officers can make decisions, determinations or examinations using automated systems made available by the Minister. This does not mean that systems themselves become the decision-maker of record. Humans design system architectures, business rules, confidence thresholds and the like, and humans are ultimately accountable for every decision the Department makes. The Minister, in turn, is accountable to Parliament for all of the Department's activities.

You will need to determine who will approve the business rules that are used in your decision support system, and whether the same person will be logged as the decision-maker of record. If an automated system determines some, but not all, elements of an administrative decision, will the person making the final determination be considered the decision-maker of record for the entire decision? Or will certain sub-elements of the overall decision be treated as separate decisions with their own decision-maker of record (e.g. eligibility vs admissibility)?

> *Add observations, analysis and links to key documents here.*

How will your system keep an audit trail for each decision?

To prepare for potential legal challenges, it is essential that IRCC establish a chain of evidence about how whole systems work, as well as a trail relating to each individual decision that goes through them.

- The automated system has the capacity to automatically generate a comprehensive audit trail of the administrative decision-making path, or you will retain an electronic record in the applicant's file
- All the key decision points are identifiable in audit trail
- All the key decision points within the automated system's logic are linked to the relevant legislation, policy or procedure
- All decisions are recorded and accessible by the system's user, a reviewer or an auditor
- The audit trail generated by the automated system can be used to help generate a notification of the decision (including a statement of reasons or other notification) where required
- The audit trail is secure from tampering (to provide protection and data integrity)
- The audit trail identifies precisely which version of an automated system was used for each decision it supports
- The audit trail includes a comprehensive and printable modification history, including: who created the record (with time and date); who has modified the record (with time and date); what was modified; for privacy and commercial-in-confidence matters, who has viewed the record (with time and date); who made the final decision (with time and date)
- The audit trail starts by identifying the authority or delegated authority identified in legislation
- The audit trail shows who the authorized decision-maker is
- The audit trail enables the recording of human intervention in automated processes, for example recording who is authorized to exercise intervention

For eTA and Visitor Record and Study Permit Extensions, the automation tab in GCMS allows users to clearly see all sub-activities, including when applications "passed" or were dropped out for manual review.

> *Add observations, analysis and links to key documents here.*

Are you assembling documentation outlining the development and deployment of the system as a whole?

This should include design specifications, training data, system versions and any other information that could be valuable in meeting TBS Directive requirements or in the event of an audit, investigation, enforcement action or judicial proceeding.

> *Add observations, analysis and links to key documents here.*

A-2021-28399-000030

How will your approach to ongoing quality assurance ensure that your automated system consistently produces the right outputs?

You may decide to start by having humans review every decision that is put through the model, and then gradually scale back the share of decisions that are human-reviewed. You should also ensure that you test for unintended fettering of officers, bias and discrimination.

In the course of a quality assurance exercise, you should make sure to document your process and its results, providing enough detail to allow others to repeat the exercise or to review it if an audit or legal challenge occurs in the future.

> *Add observations, analysis and links to key documents here.*

Security

Is your system exposed to an unacceptable level of cybersecurity risk?

Ensuring the security of environments where IRCC develops, tests and deploys AI models is critical to system integrity, compliance with privacy obligations, and public trust in the immigration system.

IRCC's IT Security Risk Management Framework aims to preserve the confidentiality, integrity and availability of departmental information and IT assets, protect the authenticity of electronic transactions, and ensure continuity of services and business operations. This framework fosters best practices, promotes the principles of continuous risk management throughout the system development lifecycle process, and lays a solid foundation for IRCC's Security Assessment and Authorization (SA&A) process.

The SA&A process (a requirement under the *Policy on Government Security*) provides a level of assurance based on the expected degree of injury that can reasonably be expected from unauthorized access and/or disclosure of data. Ultimately the SA&A provides a documented baseline of evidence of due diligence in the implementation of cybersecurity controls that supports the core principles of responsibility, privacy, chain of evidence, demonstrable transparency, and oversight.

This process should include:

- A system threat risk assessment
- A review of compliance with existing cybersecurity policies and identified security controls
- Confidence that the system – including the data it is using – is protected against hacking and manipulation

> *Add observations, analysis and links to key documents here.*

C. Preparing for launch

This section will help to answer questions like: *Is IRCC truly ready to implement the new system? Have you engaged all the right people and is the Department prepared to respond to questions and unforeseen disruptions?*

The process for getting final approval to launch may depend on the scale of changes you are proposing.

Transparency

Have you finalized your approach to public transparency?

IRCC should disclose to clients and Canadians all instances (lines of business, specific functions) in which automated systems are used are used to support decision-making. IRCC privacy, legal, technical, and communications experts will need to be consulted to help determine what information will be disclosed publicly and how it will be disclosed.

Public disclosure about systems should include:

- A plain language description of the system, including examples
  - the purpose of the system
  - the rationale for its use
  - its role within the decision-making process
  - possible outcomes for people
- A general description of the data underpinning the system and where it came from (including a link to any anonymized training data that is publicly available)
- A general description of the types of data the model uses
- Performance information, such as accuracy/error rates, distribution of affected clients (e.g. by country, by program line), cost or productivity vs humans-only alternative

Fully disclosing source code to the public is highly transparent in spirit, but most people cannot make sense of code. Thus, what can be meaningfully revealed to the general public is different from what can be shared with a community of technical experts. Where legally permissible and feasible without exposing departmental systems to vulnerabilities, IRCC should disclose source code to a panel of technical experts, and at the same time provide a plain language explanation of the system's workings to the public at large. Adopting an "Open by Design" approach has the added benefit of safeguarding the Department from cyber-attacks or sophisticated fraud on the part of malefactors with technical skills, thereby balancing transparency with program integrity. The expert panel reviewing the source code and broader system design should produce a report that includes a section suitable for public consumption.

If there are any system details that you feel should be protected from public disclosure, you will need to confirm whether this information can be protected. For example, are there suitable exemptions under the *Access to Information Act* and other legal protections available, such as for protection of national security (and, if applicable, under the TBS Directive)?

Add observations, analysis and links to key documents here.

Immigration, Refugees   Immigration, Réfugiés
and Citizenship Canada   et Citoyenneté Canada

Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

Policy Playbook on Automated Support for Decision-making – DRAFT

Have you prepared a notice for clients, alerting them to the fact that a decision on their application will be made in whole or in part by an automated system?

Notices at the point of collection of personal information should be written in plain language and easily findable on the program webpage. The TBS *Directive on Privacy Practices* and the *Directive on Automated Decision-Making* provide additional detail about the requirements for privacy notices.

The following except in the Privacy Notice Statement that appears on the TRV application form:

> *The personal information collected on an application, and other information collected in support of an application, may be used for computer analytics to support processing of applications and decision making, including your application. Personal information, including from computer analytics, may also be used for purposes including research, statistics, program and policy evaluation, internal audit, compliance, risk management, strategy development and reporting.*

For guidance around privacy notice statements, please contact the ATIP Division.

> *Add observations, analysis and links to key documents here.*

Have you consulted your communications advisor?

- Have you identified necessary updates to IRCC's external-facing page on automated decision-making?
- Will a communications strategy needed?
- Will there be an announcement?
- Have you prepared key messages, Qs&As, etc.?

> *Add observations, analysis and links to key documents here.*

Accountability

Have you worked out a governance structure that sets clear roles and responsibilities for the main partners delivering the project, including when it transitions from pilot to steady state?

The TBS *Directive on Automated Decision-Making* specifies that the Assistant Deputy Minister responsible for the program using an automated decision system is responsible for requirements related to algorithmic impact assessments, transparency measures, quality assurance, client recourse, and reporting. However, specific roles and responsibilities will vary according to IRCC's broader governance structures and individual projects.

> *Add observations, analysis and links to key documents here.*

Have you met all privacy requirements and addressed any risks that were identified in your Privacy Assessment or by the Office of the Privacy Commissioner (via your engagement with IRCC's ATIP Division)?

Engage the ATIP Division early on to make sure that measures to mitigate privacy risks can be built in during the design of your initiative, rather than at the end.

> *Add observations, analysis and links to key documents here.*

Have you met all Security Assessment and Authorization requirements and addressed any risks that were identified in your Security Action Plan (SAP)? Do you have an Authorization Decision Letter?

> *Add observations, analysis and links to key documents here.*

Do you have a strategy to ensure that project managers can track progress, measure and report on results, and support future audits or evaluations?

Ensure that you have established appropriate metrics to measure specifically the impact of the automated function(s) of your program.

> *Add observations, analysis and links to key documents here.*

Peer review

Have you consulted appropriate experts who can review the design of your system?

Depending on the potential impact of your project, the *TBS Directive on Automated Decision-Making* may require you to do this. Even if not compulsory, consultation with experts in AI or data science (e.g. from the National Research Council or Statistics Canada) may yield helpful suggestions for an improved system.

> *Add observations, analysis and links to key documents here.*

User readiness

Have you provided officers with training and instructions that are clear and user-friendly?

Understanding the strengths and limitations of the system will help users comprehend how it can best support them and how to use it appropriately. If officers don't have a high level of confidence in how an automated system works, they may be more likely to disregard the insights it generates. On the other hand, if officers place too much confidence in the system or are unsure about how to interpret results, they may tend to defer to the system, which could be seen as fettering discretion.

— Develop training materials and ask users to vet content for clarity and relevance.

— Make FAQ documents readily available to officers and other staff.
— Working with appropriate experts, consider whether Program Delivery Instructions need to be updated to explain the roles and responsibilities of employees and their use of information coming from these systems.
— Establish feedback loops to allow officers to provide input on their use of the system(s). This type of mechanism will ensure that users can share their first-hand experience and make their voices heard when challenges arise.
— Develop change management strategies, as needed. It is important not to underestimate the impact that automated systems can have on individuals, teams and our organization as a whole. The Culture and Change Management Branch can support you in identifying appropriate measures to support a smooth transition.

> *Add observations, analysis and links to key documents here.*

System readiness

Have you devised a contingency plan in the event of a system outage (short-term and long-term)?

See TBS requirements related to contingency systems and processes in the *Directive on Automated Decision-Making.*

> *Add observations, analysis and links to key documents here.*

Partner readiness

Have you confirmed with affected partners – both internal and external – that they are prepared for the launch of your system?

Consider whether partners will require any support, such as briefing materials or communications products.

> *Add observations, analysis and links to key documents here.*

Recourse

Are there sufficient recourse avenues for clients who wish to challenge the decision on their application? Does the introduction of automation necessitate any changes or additional pathways?

When a client challenges a decision that was made with the support of automation, and IRCC is authorized and agrees to review it, the review process should be such that (A) the decision reviewer stands in the shoes of the original decision-maker, and (B) the reviewer is not bound by the original decision, nor are they bound to take into account any findings or recommendations offered by the computer system that the original decision-maker relied upon. If the same technology used by the

35

A-2021-28399-000035

original decision-maker is also used by the decision reviewer, there is a risk that a flaw or error in the technology will carry over from the initial decision to the review.

Add observations, analysis and links to key documents here.

D. Once up and running

This section will help to answer questions like: *Is the system still functioning as originally intended? Do any intervening factors point to the need for a review?*

Quality assurance and monitoring

Are you conducting ongoing quality assurance to verify that your system is not generating errors or unintentional outcomes?

For example, this could include reserving a significant share of overall applications for human-only review or randomly selecting cases that go through the automated system and mixing them in with human decision-makers' normal caseload. In both cases, you would need to make sure that officers don't know which applications are new and which are quality assurance checks.

You should also monitor the operation of the system to detect and avoid administrative law and privacy risks, and determine whether any residual discriminatory impacts can be detected.

> *Add observations, analysis and links to key documents here.*

Do your results to date suggest that IRCC could safely lower the system's confidence threshold in order to realize greater efficiencies? Or the reverse?

Internal stakeholders, including Legal Services and relevant program policy areas, should be involved in reviewing results and changes to model rules.

> *Add observations, analysis and links to key documents here.*

Are you monitoring for environmental security?

This should include monitoring for new cybersecurity threats and overall system integrity.

> *Add observations, analysis and links to key documents here.*

Are you monitoring the results against the anticipated impacts you identified in previous stages of the project – e.g. in your GBA+ analysis or diagnostics?

This approach will allow you to assess whether measures to mitigate risk to diverse populations are adequate, and can help identify any additional unintended negative consequences.

> *Add observations, analysis and links to key documents here.*

Has any new data become available that could improve the next iteration of your decision support system?

If so, consider whether new data could replace proxies with more direct evidence of outcomes.

> *Add observations, analysis and links to key documents here.*

Has your model generated new insights about your program?

For instance, the expression of interest system facilitated by Express Entry has highlighted opportunities to modernize the entire federal economic program. As well, the data analytics underpinning the China TRV predictive model has bolstered the rationale for the CAN+ program (i.e. travel history).

You will need to consider questions like whether the new system will effectively replace program requirements established in legislation or regulations with algorithmically-generated business rules.

> *Add observations, analysis and links to key documents here.*

Is there an established cycle and process for periodic review of the system's business rules, data quality, confidence thresholds and productivity gains?

Without an established approach to periodic review, there is a risk that the proper functioning of a system will depend on certain key individuals staying in their same job. IRCC needs to identify limitations of algorithms (and the data underpinning them) consistently to decision-makers, regardless of staff turnover or changing roles.

IT Security Authorization Letters must be renewed at least once every 5 years, or upon significant changes to the system.

> *Add observations, analysis and links to key documents here.*

Are you assessing the overall performance of the system and reporting on results?

Depending on the objectives you established at the start, you may wish to consider how automation has helped us to reduce costs, save time, detect risks to program integrity, or improve client satisfaction.

You will also need to assess and report on how the system is contributing to meeting overall program objectives (see reporting requirements set out in the *Directive on Automated Decision-Making*).

> *Add observations, analysis and links to key documents here.*

Immigration, Refugees and Citizenship Canada
Immigration, Réfugiés et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

User insights

Are you monitoring the use of automated systems and gathering feedback from users?
As human responses to an automated system can vary over time (and the humans themselves can change as a result of staff turnover), systems should be monitored on an ongoing basis to avoid unintended negative outcomes.

— Involve users in ongoing quality assurance processes. A portion of client applications should be reserved for manual processing by officers. In addition to controlling for consistency between officer and machine decisions (e.g. through "blind testing"), this will help officers maintain the skills necessary to process applications with a range of levels of complexity.
— Leverage officers' knowledge to inform system changes and updates. Taking into account human intelligence will ensure that the system remains relevant in light of evolving country-specific risks and trends. Before scaling up or transitioning from a pilot phase to a steady state, engage a representative group of users to gather their feedback, observations and lessons learned.
— Monitor changes in users' behaviour and use of the system. You designed the system with the user in mind, you tested and tweaked machine outputs and interface, and you communicated relevant information to officers. However, it is possible that behaviour will change over time. As officers gain experience interacting with the system, they may become more, or less, trusting of machine outputs. To detect inadvertent automation bias or overreliance on system outputs, consider engaging users for periodic focus groups, observing users' behaviour in real environments, and analyzing data for anomalies. Users should also be involved any time you are considering changes that could impact the design of system outputs or interface.
— Keep users in the loop by circulating reports on the system's performance. This will demonstrate openness and transparency, and provide users with greater appreciation of the benefits (and challenges) related to automated support for decision-making.

Add observations, analysis and links to key documents here.

Transitioning from pilot to steady state

Taking your system from the pilot phase to mainstream implementation is a pivotal moment to take stock. Work with a multidisciplinary team to review questions like what worked well, what didn't, and what you would change if you make if only you knew then what you know now.

– Have you identified transitioning to steady state require additional safeguards, engagement or disclosure?
– Will your project require additional software or hardware?
– Have you tested the project for scalability?
– Have you learned lessons over the course of the pilot that could inform IRCC's future use of automated support for decision-making?

Establishing a formal reporting at the end of a pilot is not only good practice for effective project management. It also strengthens accountability and transparency, and provides greater opportunity for others to learn from your experience.

Add observations, analysis and links to key documents here.

PART III – AN OVERVIEW OF LEGAL CONSIDERATIONS AND PRACTICAL TIPS

The rule of law helps to ensure that the law is administered in a transparent and predictable manner, and provides a form of accountability to individuals affected by administrative decisions. The same legal principles and obligations of public law apply when AI is used to support administrative decision-making.

Highlighted below are some key legal considerations that arise – specifically with respect to administrative, human rights and privacy law – when using AI and automation to support administrative decision-making. These are intended to guide IRCC staff throughout the AI lifecycle, and include a set of practical tips to help mitigate risk.

Working with a variety of multidisciplinary experts – including IRCC's policy and programs, operations, data, privacy, risk, client experience, communications, and transformation branches – will also help you to identify legal risk and adopt relevant mitigation strategies throughout various stages of your project. Given that legal considerations are context specific, **you must also engage IRCC's Legal Services Unit on an ongoing basis** to seek advice pertaining to the use case at hand (this is also required under the TBS *Directive on Automated Decision-Making*).
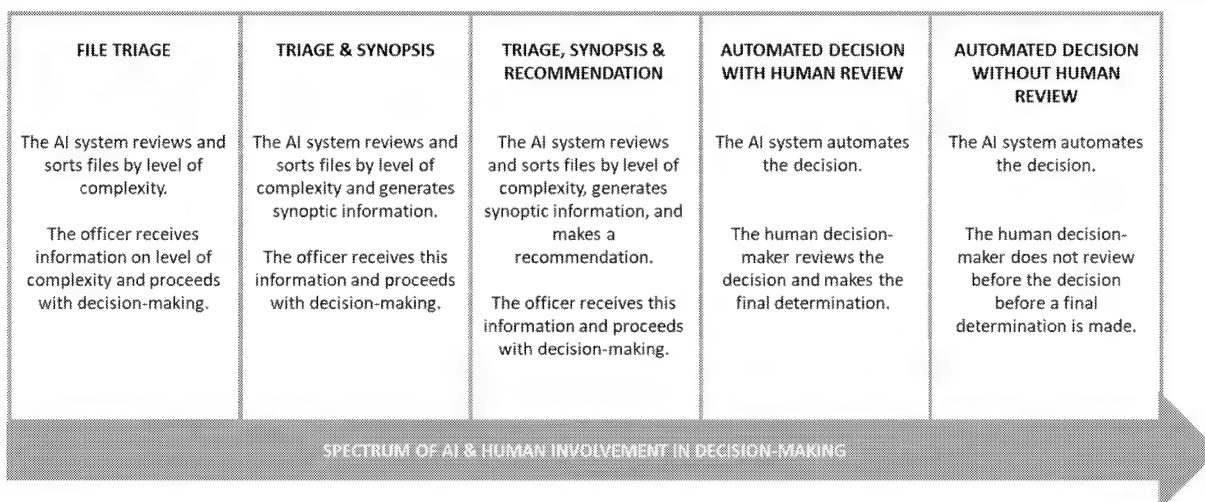
Administrative law

Administrative law accountability focuses on three essential concepts: legality, fairness, and reasonableness of decision-making. Issues that require particular attention in IRCC's environment include:

- Ensuring procedural fairness – an essential element of procedural fairness is meaningful participation. This means that the legislative criteria and the factual information used for decision-making are known to the applicant in advance, and that the individual has an opportunity to respond to it;
- Ensuring defensible decision-making – decisions are reasonable with a sufficiently transparent and intelligible justification;
- Avoiding undue or inappropriate reliance on AI – officers understand the intended scope of purpose of an AI system, and use the information provided by the system accordingly; and
- Considering the appropriate criteria and information, and mitigating bias – the information used by an AI system is consistent with legislative criteria.

A number of factors, including the degree of human involvement in design and development of AI systems and the basis for system changes over time, may impact administrative laws considerations. However, in general, the more an AI tool affects the decision-making process, the more care must be taken to ensuring that the principles of administrative law continue to be respected. A similar sliding scale of risk mitigation requirements calibrated to the level of impact of automated systems is reflected in the Government of Canada's Algorithmic Impact Assessment tool and *Directive on Automated Decision-Making*.

Immigration, Refugees Immigration, Réfugiés
and Citizenship Canada et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

FILE TRIAGE	TRIAGE & SYNOPSIS	TRIAGE, SYNOPSIS & RECOMMENDATION	AUTOMATED DECISION WITH HUMAN REVIEW	AUTOMATED DECISION WITHOUT HUMAN REVIEW
The AI system reviews and sorts files by level of complexity. The officer receives information on level of complexity and proceeds with decision-making.	The AI system reviews and sorts files by level of complexity and generates synoptic information. The officer receives this information and proceeds with decision-making.	The AI system reviews and sorts files by level of complexity, generates synoptic information, and makes a recommendation. The officer receives this information and proceeds with decision-making.	The AI system automates the decision. The human decision-maker reviews the decision and makes the final determination.	The AI system automates the decision. The human decision-maker does not review before the decision before a final determination is made.

SPECTRUM OF AI & HUMAN INVOLVEMENT IN DECISION-MAKING

Human rights law

Human rights law provides another important lens for thinking through the legal considerations that arise when using AI to support decision-making. Human rights law is relevant both because of the nature of certain administrative decisions, and because the use of AI tools can give rise to human rights considerations even in situations where the administrative decisions being made would not ordinarily affect an individual's human rights. For example, an AI system may use a client's address or geographic location information inadvertently as a proxy for national or ethnic origin, or membership in a particular religious group. Human rights protections related to equality and discrimination are of particular relevance in IRCC's context.

Two key questions to ask when developing and deploying an AI model are:

1) **Does the model create distinctions based on enumerated or analogous grounds?** Enumerated grounds are those explicitly listed under Section 15 of the Canadian Charter of Rights and Freedoms, i.e. race, national or ethnic origin, colour, religion, sex, age, or mental or physical disability. Analogous grounds are prohibited grounds of discrimination identified by the courts, e.g. non-citizenship, marital status, sexual orientation, status of being adopted, and place of birth.

2) **Are these distinctions discriminatory?** Note that not all differential treatment is discriminatory. In some cases, it may be relevant to take into account personal characteristics in light of legislative criteria.

Privacy law

Under the *Privacy Act*, accountability for government institutions focuses on principles and obligations related to the collection, notice, use, accuracy, disclosure, retention, and disposal of personal information. While the Act is designed to be technologically neutral, there may be greater potential for privacy intrusion in the context of AI and automation given the sheer volume of personal information involved.

Immigration, Refugees
and Citizenship Canada

Immigration, Réfugiés
et Citoyenneté Canada

Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

IRCC staff involved in AI projects must practice strong data and records management, and take steps to ensure that the use of personal information in AI tools is consistent with the purposes for which it was initially collected. Familiarize yourself with applicable privacy rules and requirements and consult with privacy experts at the outset to ensure privacy considerations remain front of mind during the entire AI lifecycle.

Practical tips to align with administrative, human rights and privacy law

The following practical tips suggest some steps to effectively 'put the law into practice' when approaching AI projects. They are designed to assist practitioners in moving from interpreting and applying the legal frameworks, to identifying key risks in the areas of administrative, human rights and privacy law, and mitigating these risks.

– **Consider what personal information may be used to accomplish an intended purpose, and use the information with the lowest privacy interest.** Records of model testing and any identified tradeoffs should be kept to document the rationale for decisions affecting the use of personal information.
– **Ensure notification obligations are met in relation to new collections of personal information or new uses of previously collected personal information.** IRCC can re-use personal information if the proposed use is consistent with the purpose for which it was originally collected, or with the consent of affected individuals. However, this requires careful review to establish a sufficiently direct connection between the initial purpose for which personal information was collected and the proposed purpose for an AI system.
– **Ensure data used is accurate, up to date, reliable and as complete as possible**. This should include assessing the suitability of training data for intended purposes, its reliability as a representative sample, and the accuracy of personal information in the data set over the lifecycle of its intended use. Working with IRCC's Chief Data Officer Branch and relevant program and policy leads will help ensure that historical data is suitable for purposes such as training an AI model.
– **Assess the extent to which de-identification[2] can be used.** While effective in minimizing privacy intrusion, this technique can be challenging as linking information to an identifiable individual can occur as new information is collected and technical means to defeat de-identification become more sophisticated.
– **Evaluate data flows to ensure collection, use, and disclosures of personal information meets legal and policy requirements.** Model developers and data experts will ensure a strong understanding of how the model works and which data it uses, whereas privacy experts will be able to provide advice on whether or not you need to complete a privacy impact assessment.
– **Assess training data for bias and evaluate whether the model/rules can either detect and correct, or, potentially embed or exacerbate bias**. This should be done in the early stages of design and on an ongoing basis as bias can involuntarily creep into a model over time. AI can

---

[2] De-identification refers to the process of modifying data such that it can no longer be attributed to an identifiable individual and cannot be reengineered to re-identify an individual.

Immigration, Refugees Immigration, Réfugiés
and Citizenship Canada et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

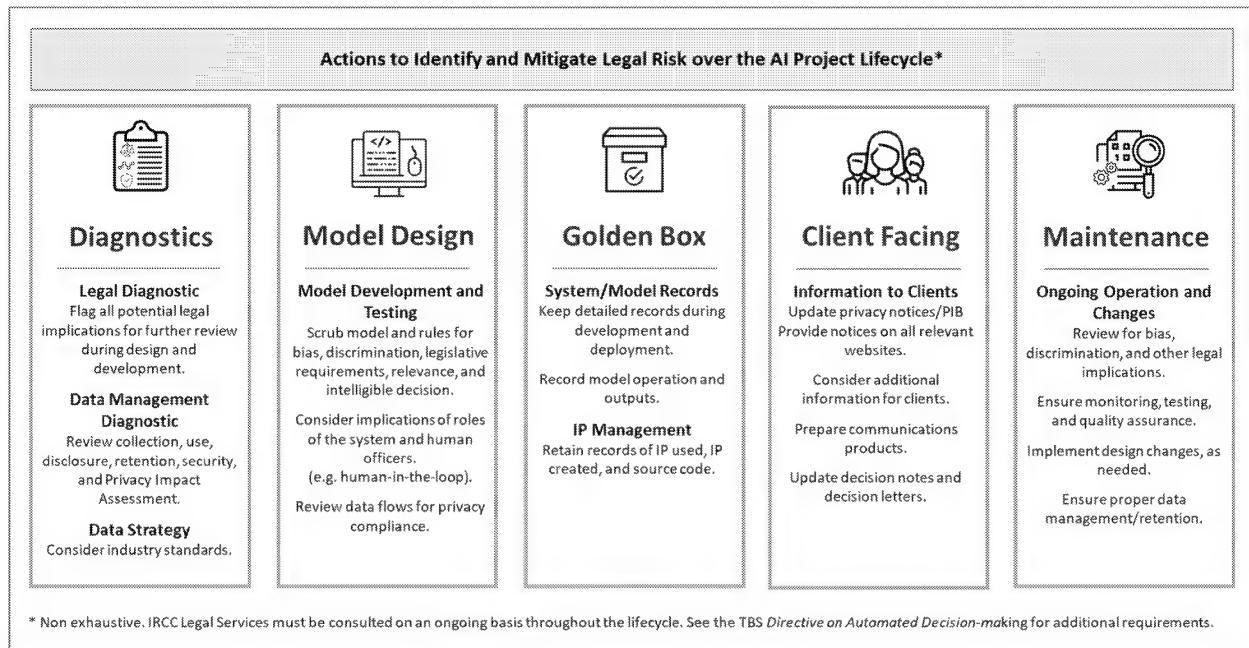*Policy Playbook on Automated Support for Decision-making – DRAFT*

be leveraged to detect and mitigate individual human cognitive bias and discrimination, enhancing the quality and consistency of decision-making.

— **Consider functionality differences with AI models and whether they impact privacy obligations.** This could involve assessing the potential impact of a privacy breach for different models or options under consideration to see how it impacts overall risk.

— **Evaluate the *algorithm* and its suitability for the intended purpose including based on best practice, industry standards (if any), and any other applicable rules or regulations.** In addition to engaging internal partners, you may need to seek advice from external stakeholders through a formal or informal peer review process.

— **Evaluate the model and rules for inclusion and consideration of legislative criteria and other relevant information**. This should include review by an experienced officer to ensure that the model and rules are consistent with traditional decision-making processes.

— **Identify whether input variables or model rules take into account enumerated or analogous grounds directly, indirectly or by proxy.** This requires a detailed understanding of the data and context, and will require collaboration with model developers, data experts, and tech counsel.

— **Determine whether it is relevant and appropriate to consider enumerated or analogous grounds, for instance, in light of legislative criteria or its relevance to decision-making.** Tech counsel, policy/program leads and experience officers can provide expertise to make informed choices in this respect. For example, the legislative criteria for becoming an economic immigrant take into account the applicant's age, but when considered alongside other factors, age is a relevant indicator of things such as the length of time the applicant is likely to remain in the workforce and their ability to adapt to life in Canada.

— **Look for patterns that may suggest differential treatment based on enumerated or analogous grounds, and assess whether this could result in discriminatory outcomes.** This assessment should be made in consultation with IRCC's tech law team, experienced officers and other multidisciplinary experts. You may need to undertake further analysis to determine the root cause (e.g. training data, algorithm, human influence, rule design) and take steps to address the issue.

— **Assess the technology and its use by decision-makers with the help of multidisciplinary experts to meet transparency and explainability requirements.** Capturing the reasons for a decision may be done in one of two ways: creating and retaining of the decision on each application, or creating the ability to reconstruct processing steps and reasons for decision after the fact.

— **Assess intelligibility and cogency of decisions and reasons available.** Depending on the context, decisions within a range of acceptable outcomes are defensible – e.g. avoiding undue reliance on AI or fettering discretion, avoiding bias, applying the legislative criteria, appropriate consideration of relevant information, and cogent reasons. AI can also be used to help decision-makers provide reasons for their decisions that are more easily understood by affected individuals, increasing accountability and potentially facilitating access to judicial review.

— **Evaluate record keeping requirements.** Information used for administrative decision-making purposes must be retained for a minimum of two years. Consideration should also be given to

Immigration, Refugees Immigration, Réfugiés
and Citizenship Canada et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

Policy Playbook on Automated Support for Decision-making – DRAFT

> keeping records in relation to 'privacy by design' choices that were made in designing and deploying AI tools.
> - **Provide adequate information and training to decision-makers** to ensure an understanding of the AI tool and the respective role of the AI tool and the human decision-maker.
> - **Monitor the operation of the AI system on an ongoing basis** to detect and avoid administrative, human rights and privacy law risks. This may include macro-level analysis to detect any remaining discriminatory impacts, or monitoring of privacy or administrative law risks that may emerge over time.

Engaging in the AI Project Lifecycle

The figure below provides a high-level snapshot of steps to be taken over the AI project lifecycle – from developing diagnostics and appropriate model design early on to providing adequate notice to clients and ensuring ongoing monitoring. Records management and retention – i.e. 'the golden box' – are also critical to the responsible use of AI in all stages of the lifecycle. This may include records of consultations, model rules and model changes, statistical analysis, quality assurance, key decisions about design and deployment, instructions provided to officers, and steps taken mitigate legal risk on an ongoing basis.

Actions to Identify and Mitigate Legal Risk over the AI Project Lifecycle*

Diagnostics
Legal Diagnostic — Flag all potential legal implications for further review during design and development.
Data Management Diagnostic — Review collection, use, disclosure, retention, security, and Privacy Impact Assessment.
Data Strategy — Consider industry standards.

Model Design
Model Development and Testing — Scrub model and rules for bias, discrimination, legislative requirements, relevance, and intelligible decision.
Consider implications of roles of the system and human officers. (e.g. human-in-the-loop).
Review data flows for privacy compliance.

Golden Box
System/Model Records — Keep detailed records during development and deployment.
Record model operation and outputs.
IP Management — Retain records of IP used, IP created, and source code.

Client Facing
Information to Clients — Update privacy notices/PIB Provide notices on all relevant websites.
Consider additional information for clients.
Prepare communications products.
Update decision notes and decision letters.

Maintenance
Ongoing Operation and Changes — Review for bias, discrimination, and other legal implications.
Ensure monitoring, testing, and quality assurance.
Implement design changes, as needed.
Ensure proper data management/retention.

* Non exhaustive. IRCC Legal Services must be consulted on an ongoing basis throughout the lifecycle. See the TBS *Directive on Automated Decision-ma*king for additional requirements.

Immigration, Refugees   Immigration, Réfugiés
and Citizenship Canada   et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

PART IV – BASELINE PRIVACY REQUIREMENTS

In Canada, privacy is considered a human right. As the majority of the data we handle at IRCC is personal information, privacy requirements must be top-of-mind when planning, developing and monitoring any initiative involving data-driven technology. Privacy implications and nuances will change based on the specifics of the initiative, however there are baseline requirements for all initiatives that must be met.

This document is intended to aid programs and policy leads in considering privacy at the outset of a disruptive technology initiative and to help guide the implementation or modification of procedures, processes and governance as may be appropriate to ensure adherence to these baseline requirements.

Throughout this document, the term 'disruptive technology' is used to encompass data science techniques and data-driven technologies such as advanced analytics, machine learning, artificial intelligence, and automation. A disruptive technology is a new technology which has a rapid and major effect on technologies that existed before[3]. Additionally, this document is meant to be read in conjunction with the Policy Playbook on Automated Support for Decision-Making.

Within this document you will find privacy-specific requirements that are mandatory regardless of the disruptive technology or automation being considered. This means that these are the **minimum requirements** and there is room for further privacy protections. The privacy requirements laid out in this document are based on the *Privacy Act*, Treasury Board Secretariat policies, directives, guidelines and internal IRCC guidance.

Following the requirements set out below is **not** a substitute for a formal Privacy Impact Assessment (PIA). Rather, these requirements will guide programs in figuring out what they have to do to be privacy compliant and may help facilitate the completion of privacy assessments as may be required. Programs intending to add disruptive technology to their programs should fill out a Privacy Needs Assessment (internal link) and send it to the ATIP Division: ATIPinternal-AIPRPinterne@cic.gc.ca.

Legal Authority

> A program must identify the parliamentary authority to collect and use personal information for the specified purposes of the program.

At IRCC, Part 4.1 of the Immigration Refugees Protection Act grants programs that fall under IRPA the ability to use electronic means to administer the Act, and section 2.2 of the Passport Order grants the Passport Program that authority. If you are unsure of your legal authority to collect and use personal information in a disruptive technology-enabled initiative, contact Legal Services.

---

[3] Adapted definition from Collins Dictionary: https://www.collinsdictionary.com/dictionary/english/disruptive-technology . Accessed on January 19, 2021.

46

A-2021-28399-000046

Immigration, Refugees Immigration, Réfugiés
and Citizenship Canada et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

Collection

> Personal information must only be collected if it relates directly to an operating program or activity of IRCC and each personal information data element must be necessary to the administration of the program[4]. When possible, personal information should be collected directly from the individual[5].

For the purposes of disruptive technology initiatives, **only information found in departmental systems of record (ex. GCMS) should be used**, and the activities of training models and algorithms should be done outside of those systems of record. Data collected from outside sources should not be used unless demonstrably necessary, and proper Information Sharing Agreements, Memoranda of Understanding, service level agreements etc. should be in place in relation to this use and strictly followed.

When considering using personal information in a model or algorithm, answer the following questions[6]:

- Demonstrating necessity: Is there a clearly defined necessity for the collection and use of the personal information? What is the problem that is expected to be solved by including this information?
- Proportionality: Is the collection and use of personal information targeted and tailored enough, so as to be viewed as reasonably proportionate to the reduction of privacy rights of the individual that the collection may represent?
- Effectiveness: Is this personal information effective at treating the issue, and so clearly connected to solving the problem?
- Minimal intrusiveness: Are the specific data elements chosen the least intrusive information available? Ensure that all other less intrusive avenues of investigation have been exhausted.
- Accuracy and validity: Is the data source a reliable and valid source of information? How will you ensure the data is accurate and up-to-date?

Notice and Informed of Purpose

> IRCC must notify individuals of the purpose for which their information is being collected, commonly referred to as a 'privacy notice'[7]. This notice must be given at or before the time of collection.

Depending on the technology being used, **it may be necessary to update existing privacy notices** and Personal Information Banks (PIB) to account for the use of the technology. If the personal information is to be used or disclosed for a purpose other than for the original purpose or a consistent purpose, a formal privacy assessment will likely be required.

---

[4] Section 4 of the *Privacy Act*.
[5] Sub-section 5(1) of the *Privacy Act*.
[6] Adapted from the Office of the Privacy Commissioner's publication *A Matter of Trust: Integrating Privacy and Public Safety in the 21st Century*. https://www.priv.gc.ca/en/privacy-topics/surveillance/police-and-public-safety/gd_sec_201011/#toc5a. Accessed January 29, 2021.
[7] Sub-section 5(2) of the *Privacy Act*.

A-2021-28399-000047

Immigration, Refugees
and Citizenship Canada
Immigration, Réfugiés
et Citoyenneté Canada

Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

Policy Playbook on Automated Support for Decision-making – DRAFT

Transparency

> IRCC must notify past applicants that their information was used to train or build models.

Updating IRCC's public-facing website about analytics and automation is one suggested way of notifying the public that past applications have informed current models. An online notice that can be modified as IRCC modifies its use of technology may also help to achieve transparency requirements.

Explainability

> Individuals have a right to know exactly how their personal information was processed through a disruptive technology system. Ensuring that plain language explanations are available on demand would allow individuals to see how technology was used to support decision-making.

Accuracy

> IRCC must take all reasonable steps to ensure that personal information used for an administrative purpose[8] is as accurate, up-to-date and complete as possible[9]. This also includes ensuring there are mechanisms in place to correct inaccurate information.

For initiatives involving disruptive technology, this involves ensuring data is collected from a reliable source, the quality of the data, developing technological mechanisms to ensure the technology is working (such as feedback loops and blind tests), quality assurance on the outputs, and so on. Additionally, to ensure the accuracy of the data, program areas must take the necessary steps to minimize unintended bias in the data. Finally, accuracy also involves model maintenance and ensuring the model is trained and re-trained on the most updated, accurate and reliable data.

Use

> Personal information must only be used for the purpose it was initially collected, a use consistent with that purpose or for a purpose for which it may be disclosed under section 8(2) (discussed under Disclosure, below)[10].

Applying disruptive technology to a dataset involving personal information is a use; this includes all uses whether administrative (as a part of a decision-making process that affects the individual) or not. Personal information must be treated appropriately regardless of the level of automation or support the technology is providing. You must also ensure that the use of disruptive technology is a consistent use of the personal information. To determine what constitutes a consistent use of personal information, the

---

[8] Annex A of the Interim Policy on Privacy Protection.
[9] Sub-section 6(2) of the *Privacy Act.*
[10] Section 7 of the *Privacy Act.*

original purpose and the proposed purpose must be so closely related that the individual would expect that the information would be used for the consistent purpose, even if the use is not spelled out. Programs that are unsure if it is a consistent use of the personal information should consult Legal Services.

The following can help to reduce the likelihood and impact of inappropriate use:

- **Data Minimization**: Only use the personal data elements that are absolutely necessary. For example: rather than full name and UCI, only the UCI is required, or rather than either, a unique identifying number can be generated for the purposes of the model only.
- **Reducing Data Granularity**: This involves removing the precision of some data elements where precision is not necessary. For example: the whole birth date might not be required, only the birth year would suffice and generate effective outputs; or the whole phone number might not be required, only the area code could suffice.
- **De-identification**: It may be possible to de-identify all or parts of datasets to reduce the likelihood of identifying individuals and reduce risks to individual privacy. Data sets may even be de-identified to the extent that it is no longer personal information which may lend to fewer data-related restrictions, yet still produce valuable results.
- **Need to Know Principle**: Only those individuals who need to know the information should have access to that information, regardless of the type of disruptive technology being applied. This may mean creating different versions of outputs: a detailed one for officers that includes a high degree of personal information, and a summary output for management that has less personal information. It may also mean limited individuals have access to the training data and where the outputs are saved. Saving information in ways that allow for modification and tailoring of access rights can help with this.

Disclosure

> Personal information under IRCC's control must not be disclosed to anyone or any organization for any reason, except for those reasons listed in sub-section 8(2) of the *Privacy Act*[11].

For initiatives involving disruptive technology, this includes information found in departmental systems of record (ex. GCMS) that other organizations such as CBSA or CSIS can view. Regular information sharing may continue to occur between IRCC and partner organizations, however Memoranda of Understanding (MOUs), Information Sharing Agreements (ISAs) and other formal agreements must be updated and modified through the appropriate channels if there is the desire to disclose the outputs on a regular basis. IRCC's <u>Admissibility and International and Intergovernmental Relations</u> (internal link) groups can provide further guidance.

Safeguards

> Personal information must be safeguarded appropriately regardless of the kind of technology applied to it.

---

[11] Section 8 of the *Privacy Act.*

A-2021-28399-000049

Immigration, Refugees   Immigration, Réfugiés
and Citizenship Canada   et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

Appropriate administrative, technical and physical safeguards should be applied to personal information for every program using personal information and therefore extends to at all stages of a disruptive technology initiative. Much like other program efforts and operations, consideration should be given to reducing the likelihood of privacy and security breaches throughout development of new technological solutions. Programs are responsible for completing the mandatory IT Security Assessment and Authorization process (internal link). Additionally, programs are responsible for developing measures to reduce the impact in the event of a privacy or security breach. IRCC's IT Security experts can offer more detailed guidance.

Retention and Disposal

> Personal information used for an administrative purpose (as a part of a decision-making process that affects the individual) must be retained for at least two years, and in accordance with the appropriate Retention and Disposition Schedule[12]. All data (with the exception of training data in the Exploration Zone) must be kept such that in the event of a complaint or legal action, the decision can be replicated.

For disruptive technology initiatives, retention and disposal decisions must be made on:
- Training data: Should be destroyed or removed from the Exploration Zone when it is no longer required for training purposes;
- Model/algorithm: Should be versioned similarly to software retention practices; and
- Outputs: Depending on the kind of model, the outputs should be added to an existing Retention and Disposition Schedule, a new RDS be created, or the outputs be deemed transitory and destroyed when no longer required.

IRCC's Information Management Branch (internal link) can provide guidance on the retention and disposition of records created around the development and deployment of disruptive technology.

Individual Access

> Individuals have a right to access the personal information that IRCC has about them[13]. IRCC should always have ways to disclose personal information on-demand.

Records around the development, deployment and monitoring of disruptive technology initiatives should also be kept at hand in the event of an *Access to Information Act* or a *Privacy Act* request.

---

[12] Sub-section 6(1) of the *Privacy Act.*
[13] Sub-section 12(1) of the *Privacy Act.*

Complaints and Reviews

> Keep clear and retrievable documents that outline all of the actions and decisions taken to ensure privacy compliance. They will be required to be produced in the event of a complaint or review.

Individuals have a right to file a formal complaint to the Office of the Privacy Commissioner (OPC) of Canada about how IRCC handles and manages their personal information, this includes how the information is processed through technology[14]. The OPC may also initiate reviews of programs when they deem necessary, and they have powers to obtain any and all information relating to the personal information handling of the program, including details about models, algorithms and automation.

Monitoring

> Monitoring for privacy compliance to the above-noted requirements should be built in from the model development phase and a monitoring schedule post-deployment should be followed that involves both the program owner(s) as well as the architects of the model.

In addition to ensuring that the disruptive technology is working properly, here is a list of non-exhaustive monitoring activities to plan for:

- **Collection**: Make sure no data from sources other than departmental systems of record (ex. GCMS) and other IRCC data repositories are included in the disruptive technology, and if there is outside data, ensure the appropriate ISA or MOU is in place and up to date.
- **Notice**: Review privacy notices and transparency and explainability practices to ensure accuracy and up to date information. Update whenever required.
- **Retention and Disposal**: Review the retention and disposal practices and ensure no information is retained for longer than it should be.
- **Accuracy**: Build in regular data quality practices to ensure data is accurate, up to date and complete as possible, and modify the information when required.
- **Use**: Review data handling practices, ensure mitigation measures against inappropriate use are functioning properly, employ new measures as required, and update practices periodically.
- **Disclosure**: Review disclosure practices to ensure disclosures are occurring as a part of up to date MOUs and/or ISAs, and ensure other government organizations can only see what information they are allowed to see in GCMS. Make modifications when required.
- **Safeguards**: Conduct regular security checks to ensure the training data, the technology itself and the outputs are secure.

This document provides an overview of the baseline privacy requirements that all disruptive technology initiatives at IRCC must consider. Additional requirements may be required based on specific initiatives, and even if you follow these requirements, a privacy assessment may be required. A multi-stakeholder approach is the best way to ensure all requirements and risks are addressed and mitigated

---

[14] Section 29 of the *Privacy Act.*

appropriately. The ATIP team and others throughout the department as referenced within this document are here to help you navigate the privacy considerations to ensure that programs are built with privacy in mind. Please fill out a Privacy Needs Assessment (internal link) to begin your consultation with the ATIP Division.

PART V – CHECKLIST FOR THE DIRECTIVE ON AUTOMATED DECISION-MAKING

The *Directive on Automate Decision-Making* is the Government of Canada's policy that governs the use of artificial intelligence to support administrative decisions. The Directive came into force in April 2020, and applies to automated systems, tools and models used by federal institutions to make or support decisions affecting external clients.

The following checklist provides an overview of the requirements you must address at various stages to comply with the Directive. You should think of these as minimum requirements; other requirements specific to IRCC's legal framework, policies and programs may also apply.

Some requirements apply to all systems that fall within the scope of the Directive, whereas others vary according to the level of impact associated with algorithm used. In addition, although the Directive only applies to automated systems in production (therefore excluding those in a testing environment), reviewing the requirements and mandatory Algorithmic Impact Assessment tool early on will help you anticipate and develop critical risk mitigation strategies in the design and testing phase of your project.

A. Exploring automated support for decision-making as a potential solution

☐ **Legal** – Have you consulted with IRCC's Legal Services? This must be done starting from the concept stage and continue on an ongoing basis.

B. Designing your system

☐ **Algorithm Impact Assessment (AIA)** – Have you completed a preliminary AIA and identified requirements related to the relevant the level of impact?

☐ **Access to components** – If applicable, have you determined the appropriate license for software components and, if using a proprietary license, addressed related requirements?

☐ **Data quality** – Have you validated that the data collected for, and used by, the automated decision system is relevant, accurate, up-to-date, and in accordance with the *Policy on Information Management* and the *Privacy Act*?

☐ **Security** – Have you assessed security risks related to the system and applied appropriate safeguards, as per the *Policy on Government Security*?

☐ **Human intervention** – Have you ensured that the system allows for human intervention? (*Levels III and IV only*)

C. Preparing for launch

☐ **AIA** – Have you finalized a full response to all questions in the AIA and taken steps to address all requirements relevant to the level of impact?

☐ **AIA** – Have you released the final version of the AIA in an accessible format?

☐ **Source code** – Have you released custom source code owned by the Government of Canada, as required? You will also need to determine the appropriate access restrictions to the released source code.

Immigration, Refugees   Immigration, Réfugiés
and Citizenship Canada   et Citoyenneté Canada
Information disclosed under the Access to Information Act
L'information divulguée en vertu de la loi sur l'accès à l'information

Policy Playbook on Automated Support for Decision-making – DRAFT

☐ **Testing and monitoring outcomes** – Have you developed processes to test data and information used by the automated decision system for unintended data biases and other factors that may result in unfair outcomes?

☐ **Testing and monitoring outcomes** – Have you developed processes to monitor outcomes on a scheduled basis, including monitoring for unintentional outcomes and verifying compliance with relevant legislation and this Directive?

☐ **Notice** – Have you provided notice on all service-delivery channels in use to inform applicants that the decision rendered will be undertaken in whole or in part by an automated system? NB: Notice must be provided prominently and in plain language; IRCC's Legal Services Unit, ATIP Division and Communications teams can provide advice on. (*Levels II-IV only*)

☐ **Peer review** – Have you consulted the appropriate qualified experts to review the system? (*Levels II-IV only*)

☐ **Employee training** – Have you developed adequate training for employees and put in place measures/a strategy to ensure employees receive the training? (*Levels II-IV only*)

☐ **Contingency** – Have you established contingency systems and/or processes? (*Levels III and IV only*)

☐ **Approval** – Have you obtained the appropriate level of approvals for the system to operate? (*This requirement applies largely to Level III and IV systems, however, the Assistant Deputy Minister responsible for the program using the automated system must ensure that all relevant requirements set out in the Directive are met.*)

D. Once up and running

☐ **AIA** – Have you updated the AIA to reflect changes in system functionality or scope? In the event of a change in level of impact, you may need to address additional requirements.

☐ **Explanations** – Have you provided a meaningful explanation to affected individuals of how and why the decision was made?

☐ **Results** – Are you storing the results of decisions, as required, and complying with monitoring and reporting requirements?

☐ **Testing and monitoring outcomes** – Is data used by the system routinely tested to ensure that it is still relevant, accurate, and up-to-date?

☐ **Recourse** – Have you provided clients with appropriate recourse options to challenge the automated administrative decision?

☐ **Reporting** – Have you published information about the effectiveness and efficiency of the automated decision system in meeting program objectives?

GLOSSARY OF KEY TERMS

Administrative decision-making: Decisions made by authorized officials of an institution (such as a board, tribunal, commission, government department, agency or Minister), where the outcome affects legal rights, privileges or interests. Central to the immigration system, administrative decision-making is also critical in other fields like international trade, taxation, broadcasting and transport.

Algorithm: A sequence of instructions, rules, and calculations executed by a computer in a particular order to yield a result, typically an answer to a specified problem. Multiple algorithms can be used in combination to solve complex problems.

Artificial intelligence: Encompassing a broad range of technologies and approaches, AI is essentially the field of computer science dedicated to solving cognitive problems commonly associated with human intelligence, such as learning, problem solving, and pattern recognition.

Automated decision support system: Includes any information technology designed to directly support a human decision-maker on an administrative decision (for example, by providing a recommendation), and/or designed to make an administrative decision in lieu of a human decision-maker. This includes systems like eTA or Visitor Record and Study Permit Extension automation in GCMS.

Big data: A popular colloquial term used to describe immense information assets that inundate organizations on a day-to-day basis. This data is often highly varied and needs to be processed through computer systems at a fast pace, making it impossible to effectively manage and yield business value without specific software tools and analytical methods.

Black Box: Opaque software tools working outside the scope of meaningful scrutiny and accountability. Usually deep learning systems. Their behaviour can be difficult to interpret and explain, raising concerns over explainability, transparency, and human control.

Business rules: A series of instructions (if-then propositions), from which an automated system is unable to deviate. Rule sets can be human-crafted (based on experience, common sense, research) or human-curated (proposed by advanced analytics models but reviewed and approved by people). Automated systems comprise business rules that cannot be modified without human intervention. This is distinct from *autonomous* systems, which have pre-determined goals but no pre-determined approaches to how they are achieved.

Data analytics (including advanced analytics): The pursuit of extracting meaning from large volumes of (sometimes high-velocity) data using specialized computer systems. These systems transform, organize, and model the data to draw conclusions and identify patterns.

Data governance: Efforts at a corporate level to formally manage data assets (e.g. who does what, when, and under what authority). Effective data governance ensures that data is of high-quality as it is created, extracted and manipulated for reporting and decision-making. A bridge between the business and IT spheres, data governance supports process innovation and improved client service.

Exploration zone: The exploration zone – also referred to as a "sandbox" – is the environment used for research, experimentation and testing related to advanced analytics and AI. Data, codes and software are isolated from those in production so that they can be tested securely.

"Fettering" of a decision-maker's discretion: Fettering occurs when a decision-maker does not genuinely exercise independent judgment in a matter. This can occur when a decision-maker binds him/herself to a fixed rule of policy, another person's opinion, or the outputs of a decision support system. Although an administrative decision-maker may properly be influenced by policy considerations and other factors, he or she must put his or her mind to the specific circumstances of the case and not focus blindly on one input (e.g. a risk score provided by an algorithmic system) to the exclusion of other relevant factors.

Procedural fairness: Crucial to administrative decision-making, procedural fairness requires that applicants:

— have meaningful participation (they know the case to be met and have an opportunity to meet it) and are provided with a fair and unbiased assessment of their application
— are informed of negative evidence and the decision-maker's concerns
— have a meaningful opportunity to provide a response to such evidence and concerns about their application
— are provided reasons for the decision

There are numerous factors that determine what, specifically, should be done to address these principles. One such factor is the importance of the decision from the applicant's perspective.

Process automation: Also called "business automation" (and sometimes even "digital transformation"), process automation is the use of digital technology to perform routine business processes in a workflow. Process automation can streamline a business for simplicity and improve productivity by taking mundane repetitive tasks from humans and giving them to machines that can do them faster. A wide variety of activities can be automated, or more often, partially automated, with human intervention maintained at strategic points within workflows. In the domain of administrative decision-making at IRCC, "process automation" is used in contrast with "automated decision support," the former referring to straightforward administrative tasks and the latter reserved for activities involving some degree of judgment.

Machine learning: A sub-category of artificial intelligence, machine learning refers to algorithms and statistical models that learn and improve from examples, data, and experience, rather than following pre-programmed rules. Machine learning systems effectively perform a specific task without using explicit instructions, relying on models and inference instead.

Source code: Computer program in its original programming language, human readable, before translation into object code. It consists of algorithms and computer instructions, and may include developer's comments.

SOUTIEN AUX DÉCISIONS AUTOMATISÉ
Guide de politique

Assurer l'utilisation responsable de l'automatisation, de l'analytique avancée et de l'intelligence artificielle dans la prise de décisions administratives

1. PRÉAMBULE

Pourquoi avons-nous besoin d'une orientation stratégique dans ce domaine?

À Immigration, Réfugiés et Citoyenneté Canada (IRCC), nous connaissons déjà la prémisse entourant l'automatisation de certains aspects de la prise de décisions administratives : le volume de demandes ne cesse de croître, notre réseau opérationnel subit une pression de plus en plus élevée, et la technologie semble être notre meilleur espoir de suivre le rythme tout en maintenant un service de qualité. Les modèles opérationnels traditionnels nécessitent simplement trop de ressources pour la portée des opérations que nous prévoyons pour 2019 et les années subséquentes.

Les domaines des données massives et de l'intelligence artificielle (IA) sont en pleine expansion, et IRCC reconnaît désormais leur immense potentiel pour l'aider à mener ses activités autrement et réaliser des gains d'efficacité. Dans cette nouvelle ère, toutefois, les méthodologies, mécanismes de protection et modèles de gouvernance sont souvent incompatibles avec le fonctionnement des approches algorithmiques. Jusqu'à ce jour, IRCC a relativement bien sorti son épingle du jeu grâce à une approche provisoire pour résoudre les questions de protection de la vie privée, de transparence, d'équité procédurale et d'interactions humain-machine. Cependant, l'IA et l'automatisation (et les risques qui s'y rattachent) prenant une place de plus en plus importante, le Ministère bénéficierait d'une approche réfléchie. Pour réussir, les innovateurs d'IRCC ont besoin d'une orientation bien définie et accessible qui soit adaptée à notre contexte opérationnel.

Le Secrétariat du Conseil du Trésor (SCT) propose une nouvelle Directive sur la prise de décisions automatisée applicable à l'échelle du gouvernement. Cette politique du SCT, qui entrera en vigueur en avril 2020, énoncera les exigences générales s'appliquant aux ministères utilisant des systèmes, outils et modèles statistiques pour recommander ou prendre des décisions administratives à propos de leurs clients. Elle sera accompagnée d'un outil de soutien aux évaluations des répercussions algorithmiques, qui deviendra obligatoire pour quiconque prévoyant faire passer un système d'un environnement d'essai à sa mise en œuvre. Ces instruments stratégiques globaux permettront à IRCC de découvrir ce terrain inconnu, mais l'orientation sera probablement trop vague pour s'attaquer à des difficultés précises que doivent surmonter les concepteurs de systèmes d'IRCC.

De plus, la Directive du SCT suggère un point de départ différent de la politique interne énoncée aux présentes. Selon le SCT, le lecteur a déjà décidé d'adopter une solution algorithmique et en est à la phase de développement du système visé. Le politique permet de s'assurer que le système, une fois terminé, respecte les normes. En revanche, la politique interne d'IRCC repart de zéro et pose des questions fondamentales, notamment : *Dans mon cas, le soutien à une prise de décisions automatisée est-il une bonne idée? Qu'est-ce qu'il faut automatiser exactement? Par où dois-je commencer?*

À titre de fournisseur de services pour les citoyens canadiens et étrangers, IRCC a besoin d'une orientation qui tienne compte de sa position particulière. Son personnel doit pouvoir s'appuyer sur des politiques qui intègrent des concepts d'éthique liés aux technologies dans ses activités quotidiennes et qui répondent directement aux questions qu'il se pose. Par exemple : *Pourrions-nous utiliser des algorithmes pour automatiser des décisions défavorables aux demandes de clients? Dans quels cas les « boîtes noires » d'algorithmes sont-elles appropriées, ou ne le sont-elles jamais? Qui est responsable des décisions individuelles prises par une machine? Comment explique-t-on clairement les décisions automatisées aux clients? Comment pouvons-nous nous assurer que les agents d'IRCC profitent des observations tirées d'analyses sans entraver leur prise de décisions?*

IRCC doit trouver réponse à ces questions. Il est essentiel d'avoir la confiance du public pour assurer la réussite des programmes d'immigration, d'asile, de citoyenneté, d'établissement et de passeport, et à l'heure actuelle, le public s'inquiète des répercussions potentielles de l'IA et de l'automatisation qui se propagent de plus en plus. Si l'on codifie ou « ancre » les règles opérationnelles, que ce soit par l'analytique avancée ou le raisonnement heuristique qu'appliquent depuis longtemps les agents, elles font alors l'objet d'une surveillance accrue. L'application de règles opérationnelles à une très grande échelle multiplie les risques d'incidents. Une nouvelle portée des répercussions amène son lot de risques juridiques pour IRCC, l'un des ministères fédéraux faisant le plus l'objet de litiges, lui qui s'appuie déjà sur des systèmes automatisés pour exécuter ses programmes. Le traitement informatique est essentiel dans le cadre de programmes tels

que l'autorisation de voyage électronique (AVE) et Entrée express, et le soutien aux décisions automatisé pourrait bientôt devenir une composante indispensable du programme de résidence temporaire et d'autres secteurs de programme.

En plus de contrôler les risques prévus, une politique propre au soutien aux décisions automatisé peut aider IRCC à réaliser tous les avantages d'une nouvelle approche. Les principes s'appliquent aux mesures que devrait éviter le Ministère en raison de leurs risques trop élevés, mais aussi aux activités que nous devrions mener parce qu'elles nous permettent de *nous améliorer*.

Quelle est la portée des activités visées par le présent guide?

La présente politique vise **les systèmes automatisés qui exécutent ou appuient directement, en tout ou en partie, la prise de décisions administratives.** Il s'agit entre autres des systèmes qui : classifient les dossiers selon le niveau de surveillance requis; signalent les dossiers qui devraient faire l'objet d'un examen ou d'une enquête par un agent; formulent des recommandations quant à l'acceptation ou au refus d'une demande; rendent des décisions du début à la fin du processus. Les règles opérationnelles qu'applique un système de soutien aux décisions automatisé pourraient découler d'analyses de données sophistiquées ou d'entrevues avec des agents chevronnés; ce que nous savons, c'est que ce système joue aujourd'hui un nouveau rôle dans le modèle de prise de décisions d'IRCC. Les règles fondées sur l'analyse sont particulièrement préoccupantes selon la présente politique, surtout si l'apprentissage machine entre en ligne de compte.

La politique n'aborde pas tous les aspects de l'automatisation, de l'IA ou des décisions administratives. En effet, elle ne porte pas sur l'automatisation de processus courants, comme la programmation d'ordinateurs pour qu'ils exécutent des tâches de bureau répétitives qui étaient auparavant accomplies manuellement. Elle ne vise pas non plus les systèmes qui aident simplement à la répartition de la charge de travail entre les réseaux, les bureaux et les agents par mesure d'efficacité. Si un processus d'automatisation ne joue pas un certain rôle pour aider IRCC à prendre une décision administrative ayant une incidence sur l'avenir d'un client, alors la présente politique n'en tient pas compte. De la même manière, elle ne vise pas les applications d'IA, quel que soit leur degré de sophistication, qui servent à améliorer le service aux clients (comme un robot de clavardage) ou aident le Ministère à gérer les ressources internes. Elle vise la prise de décisions, plus précisément les décisions qui s'inscrivent en dehors du modèle traditionnel où un agent examine un seul ensemble de faits portant sur une seule demande.

Quelle forme prend le guide?

Pour établir un continuum complet de soutien axé sur l'éthique, il faut à la fois un code d'éthique personnel et un ensemble de principes d'éthique institutionnels. Sur le plan personnel, il n'est pas nécessaire de créer de nouveaux instruments. Le Code de valeurs et d'éthique du secteur public en place s'applique à tous les domaines (il s'applique d'ailleurs déjà tout autant aux comptables menant des vérifications internes qu'aux scientifiques de données qui mettent au point des systèmes d'apprentissage machine). Les cinq valeurs essentielles du Code, soit le respect de la démocratie, le respect envers les personnes, l'intégrité, l'intendance et l'excellence, jettent des bases solides et durables servant à aider les innovateurs d'IRCC à prendre les bonnes décisions pour agir de manière éthique et professionnelle. En ce qui a trait à l'éthique institutionnelle, la partie 2 propose un ensemble de principes directeurs.

Les principes énoncés dans la présente politique visent à accompagner l'utilisateur dans la partie 3 plus concrète proposant des directives. En règle générale, les principes ne dictent pas des mesures précises, mais servent plutôt de fondement pour aider les personnes à prendre des décisions lorsqu'elles sont dans un nouveau milieu et font face à des problèmes inconnus. En diffusant les principes directeurs présentés dans la partie 2, IRCC espère inspirer, former et outiller ses scientifiques de données, concepteurs de programmes et créateurs de politiques de sorte qu'ils accordent la priorité à l'éthique dans le développement des systèmes de soutien aux décisions automatisé. Particulièrement, ils orientent le Ministère dans son ensemble ainsi que les membres de sa direction, qui sont en fin de compte responsables des résultats des systèmes automatisés et de la création d'un milieu de travail qui valorise l'utilisation responsable de la technologie. En tenant compte et en discutant des répercussions générales des efforts novateurs du Ministère, nous ne pouvons que nous assurer que ce dernier appuie de manière constante le bien du public.

La partie 3 présente un manuel pratique destiné aux spécialistes des politiques et des programmes qui envisagent de développer un nouveau système de soutien à la prise de décisions automatisée. Cette section oriente systématiquement

2

les innovateurs grâce à un processus linéaire qui les encourage à poser les bonnes questions au bon moment : (1) quand ils doivent déterminer si une solution fondée sur des données convient au problème à régler, (2) quand ils prévoient créer un nouveau système, (3) quand ils se préparent au lancement, et (4) quand le système automatisé est fonctionnel.

À la fin du document figure un glossaire des termes clés.

> *Une nouvelle politique sur le soutien aux décisions automatisé constitue pour le Ministère une occasion de s'assurer que sa réflexion suit le rythme du changement technologique et que ses gens et ses pratiques continuent d'offrir une gamme de programmes à la hauteur des attentes des Canadiens et du monde entier. Les normes et les principes fondamentaux qui guident IRCC demeurent inchangés. Nos engagements, notre mission et nos devoirs restent les mêmes. Même s'il peut sembler que les outils que nous utilisons à l'ère numérique reflètent simplement les pratiques en personne et sur papier d'autrefois, il n'en est rien. L'adoption de nouvelles technologies a changé et changera ce que nous faisons et comment nous le faisons, les compétences qu'il nous faut, nos relations avec nos partenaires et nos clients et notre culture en milieu de travail.*

<u>Remarque</u> : Veuillez considérer la présente politique comme un document évolutif qui doit être régulièrement mis à jour pour demeurer pertinent, à mesure qu'évoluent les technologies et nos connaissances à cet égard.

2. UN ENSEMBLE DE PRINCIPES DIRECTEURS

Comment le Ministère dans son ensemble devrait-il agir face à des choix difficiles sur certaines applications de technologies émergentes? Nous proposons un ensemble de principes pour aider IRCC à faire les meilleurs choix sur les plans de l'efficacité, de la reddition de comptes et de la conception responsable.

PRINCIPAUX OBJECTIFS

1. *L'utilisation de nouveaux outils devrait s'accompagner d'avantages évidents pour le public. IRCC devrait utiliser un soutien aux décisions automatisé quand il peut le faire de manière responsable, efficace et efficiente, dans cet ordre.*

❖ **En utilisant le soutien aux décisions automatisé, IRCC devrait renforcer la confiance du public à l'égard du système d'immigration du Canada**. Le maintien de la confiance des Canadiens est essentiel à la réussite du Ministère et ces derniers se montrent méfiants quant aux risques touchant l'équité et la vie privée associés aux données massives et à l'IA. Le Ministère devrait suivre des étapes calculées, particulièrement au début, pour asseoir la légitimité de son utilisation d'outils d'automatisation, d'analytique et d'IA. Il faudrait concevoir et mettre en œuvre les nouveaux outils de manière à inspirer confiance et à favoriser la compréhension.

❖ Il existe un coût de renonciation associé au fait de <u>ne pas</u> adopter l'automatisation, l'analytique avancée et l'IA. IRCC ne devrait pas faire preuve de prudence excessive pour ne pas manquer cette occasion au potentiel énorme que représentent les technologies fondées sur les données. Comme dans bien des domaines, **l'utilisation des meilleurs outils à notre disposition fait partie d'un usage efficace et efficient de l'argent des contribuables**. Dans certains cas, les systèmes algorithmiques sont le seul moyen efficient de traiter l'énorme volume d'information requis pour la prestation de services gouvernementaux modernes.

3

❖ Cependant, **IRCC peut seulement utiliser ces technologies dans la mesure où leur utilisation est conforme aux normes canadiennes** en matière d'égalité et de non-discrimination, d'équité procédurale, de transparence, de respect de la vie privée et de responsabilisation. La technologie ne change aucunement les exigences du droit administratif et des cadres de travail généraux comme la *Charte canadienne des droits et libertés* et la *Loi sur la protection des renseignements personnels.*

❖ **IRCC ne devrait pas adopter l'approche de décisions automatisées (ou certains aspects des décisions) au détriment de l'intégrité des programmes.** Le travail d'IRCC repose sur la gestion du risque. L'automatisation peut changer les risques ou en introduire de nouveaux, mais l'innovation devrait améliorer l'intégrité des programmes. Au pire, l'incidence devrait être neutre à cet égard.

❖ Les données massives et l'automatisation devraient **renforcer la prise de décisions administratives en fournissant une foule de renseignements pertinents aux décideurs et en orientant leur attention vers les facteurs déterminants**. De cette façon, les nouveaux outils devraient aider les agents à prendre des décisions plus rapidement, mais surtout à prendre de *meilleures* décisions.

❖ **IRCC devrait déployer tous les efforts possibles pour obtenir des données fiables sur les résultats** (par exemple d'un système d'intrants et d'extrants, de déclarations fiscales, des provinces et territoires) pour faire un lien direct entre les résultats finaux des clients et l'information qu'ils soumettent à l'étape de la demande. Simultanément, le **Ministère doit dûment tenir compte de l'impact des conclusions qu'il tire des nouveaux ensembles de données**. Faire la distinction entre la corrélation et la causalité est un défi constant et il est difficile de prévoir avec exactitude les intentions des humains.

❖ **Les données massives devraient aider IRCC à se tourner vers le passé comme vers l'avenir.** Tandis que des algorithmes mal conçus peuvent introduire ou perpétuer des préjugés à grande échelle, des **modèles bien conçus peuvent faire la lumière sur des préjugés historiques** en décelant des tendances de données qui étaient jusque-là encore inconnues. IRCC devrait envisager d'utiliser de nouveaux outils pour détecter, évaluer et éliminer tout préjugé ou toute incohérence qui pourraient s'intégrer par inadvertance dans les pratiques de longue date.

❖ **IRCC devrait chercher à améliorer continuellement la qualité de ses fonds de données, puisque l'analytique avancée et l'IA reposent sur des données de qualité.** En accordant la priorité au développement d'outils fondés sur les données, nous accordons aussi la priorité à la portée, à la profondeur et à la fiabilité de nos ensembles de données.

2. *Les décisions administratives concernent des gens et sont prises par des gens, même si l'IA entre en jeu. Ce sont des humains, et non des systèmes informatiques, qui sont responsables des décisions.*

❖ **L'adoption d'un soutien aux décisions automatisé peut modifier** *le moment et l'endroit* **où l'humain intervient dans le processus de prise de décisions** (par exemple, l'humain peut jouer un nouveau rôle, comme établir des règles opérationnelles pour un système de triage automatisé à déployer plus tard), **mais ce mécanisme de soutien n'enlève pas à l'humain son jugement dans la prise de décision**. Les humains jouent un rôle fondamental dans la décision concernant les types de systèmes à utiliser, les dossiers auxquels ces systèmes s'appliquent et les valeurs à programmer dans ces systèmes.

❖ **Il est important qu'un humain ait sa place dans le processus pour inspirer confiance au public. De nouveaux modèles opérationnels qui intègrent un soutien aux décisions automatisé devraient, d'une manière ou d'une autre, prévoir une place pour les humains dans le processus de prise de décisions.** Bien que la prise de décisions

4

par les humains puisse ne pas être supérieure à des systèmes algorithmiques reposant sur une analytique de données sophistiquée dans tous les cas, des systèmes comportant l'intervention humaine dans son processus représentent actuellement une forme de transparence et de responsabilité personnelle qui est mieux connue du public que les processus automatisés. Dans les cas où des algorithmes contribuent à la prise de décisions administratives qui touchent considérablement la vie des gens (comme le fait d'accorder ou non à une personne sa résidence permanente), il est raisonnable que les demandeurs et les Canadiens s'attendent à ce qu'un humain ait utilisé son jugement pour développer le processus et que quelqu'un soit responsable des résultats finaux.

❖ L'automatisation, l'analytique avancée et l'IA sont des multiplicateurs de force; elles **appuient et accroissent les capacités des gens**. Dans le contexte des volumes d'information entrante en constante croissance dans tous les secteurs d'activité d'IRCC, la technologie devrait aider notre effectif à simplifier ses efforts. En permettant au personnel d'IRCC de se concentrer sur des tâches axées sur la créativité, la résolution de problèmes ou l'interaction avec les clients, nous améliorerons leur prestation de service et leur satisfaction au travail.

❖ **IRCC devrait s'efforcer de fournir à ses décideurs la meilleure information et les meilleurs outils disponibles.** En mettant à la disposition des décideurs de nouvelles connaissances et en leur permettant de centrer leur attention sur les facteurs les plus importants, nous éliminerons les efforts inutiles, améliorerons l'uniformité et lierons mieux l'action et sa justification.

❖ Les machines ne se voient pas transférer les responsabilités lorsqu'elles exécutent des tâches auparavant accomplies par des humains. **IRCC doit assumer la responsabilité des réussites et des échecs de ses systèmes.** En disant que les systèmes ou les outils sont « biaisés », on transfère la responsabilité des résultats négatifs à l'algorithme. C'est comme si l'on confiait en sous-traitance son obligation morale.

LES BONS OUTILS DANS LES BONNES CIRCONSTANCES

3. *Puisque les décisions d'IRCC ont une incidence importante sur la vie de ses clients et des Canadiens, le Ministère doit accorder la priorité aux approches qui comportent le risque le plus faible.*

❖ L'automatisation des décisions définitives constitue l'approche la plus risquée pour réaliser des gains d'efficacité. Pour cette raison, et parce que les décisions définitives sont rarement la portion la plus exigeante sur le plan des ressources de la prise de décisions, IRCC devrait d'abord **examiner s'il peut réaliser les mêmes gains d'efficacité par l'automatisation d'autres aspects du processus opérationnel** (p. ex. la répartition de la charge de travail, le triage en fonction des risques, la prise de notes, la recherche et les communications). L'automatisation des processus, comme la vérification de l'état complet des demandes, peut préparer la voie pour les décideurs, afin qu'ils exercent leur jugement sans avoir d'abord à surmonter des obstacles que les machines peuvent éliminer.

❖ **Le degré d'intervention humaine dans la prise de décisions devrait être calibré en fonction des critères suivants :**
    1. la gravité de la décision, y compris son incidence sur les clients;
    2. la complexité et le degré de jugement subjectif nécessaire;
    3. la qualité des données qui peuvent être prises en compte.

Ces considérations sont plus importantes que de savoir si l'action en question est une décision définitive ou une étape du processus ou s'il s'agit d'une décision favorable (approbation) ou défavorable (refus).

❖ Dans le cas d'un modèle prévisionnel pour la prise de décisions, on considère que l'automatisation des approbations a habituellement moins d'incidence que l'automatisation des refus (étant donné que cette dernière opération peut être perçue comme un profilage injuste). C'est généralement vrai du point de vue d'un demandeur individuel. Cependant, il est important de voir la situation de la perspective d'autres demandeurs, particulièrement ceux qui font une demande dans un secteur d'activité où le nombre de places est limité. **Dans les situations où la demande est supérieure à l'offre, les risques des décisions favorables sont plus élevés** : cela peut être perçu comme si l'approbation (ou encore l'invitation ou l'acceptation) d'une demande d'un candidat est accordée au détriment de celle d'un autre.

❖ **Une troisième perspective est celle de l'ensemble des Canadiens**, qui profitent des avantages économiques, sociaux et culturels de la migration temporaire et permanente, mais qui aussi s'exposent aux risques de sécurité connexes. **De la perspective collective, les approbations automatisées fondées sur la prévision peuvent avoir des conséquences importantes.** Un système qui approuve des demandes sans validation suffisante augmenterait les risques pour les Canadiens, et il est compréhensible que ces derniers s'inquiètent davantage de l'approbation par erreur de la demande d'une personne à risque que le refus par erreur de celle d'un candidat de bonne foi. IRCC doit adopter un point de vue équilibré, et tenir compte de l'équité de la perspective individuelle et de la perspective collective.

4. *Les algorithmes « boîtes noires » peuvent être utiles, mais ne peuvent pas être le seul déterminant dans les décisions définitives concernant les demandes de clients.*

❖ **Les véritables outils d'IA « boîtes noires » sont d'une utilité limitée dans la prise de décisions administratives**, car le Ministère ne devrait pas prendre des décisions qu'il ne peut expliquer clairement. Les outils « boîtes noires », comme la reconnaissance faciale, peuvent jouer un rôle de soutien. **Toutefois, même dans ce cas, le degré de confiance et de fiabilité devrait être considérablement élevé, et sinon, des humains devraient pouvoir examiner les résultats du système avant la prise de décisions définitives**.

❖ Ce ne sont pas toutes les décisions d'IRCC qui touchent les droits fondamentaux des clients et qu'il faut expliquer à ces derniers. **Les algorithmes « boîtes noires » devraient être permis aux fins d'optimisation opérationnelle**, par exemple pour définir l'emplacement des bureaux ou comment mieux répartir les ressources entre eux.

CONCEPTION RESPONSABLE

5. *IRCC reconnaît les limites des technologies fondées sur les données et prend toutes les mesures raisonnables pour réduire au minimum les préjugés involontaires.*

❖ Les données portent toujours les marques de l'histoire. Si nous utilisons les données pour former un système afin qu'il formule des recommandations ou prenne des décisions, nous devons être pleinement conscients du fonctionnement de l'histoire. Chaque algorithme d'apprentissage machine fonctionne entièrement dans un monde défini par les données utilisées pour le calibrer; ainsi, **les limites ou les défaillances dans les ensembles de données nuisent aux résultats**, parfois gravement. Au moment de créer des outils pour la prise de décisions concernant les demandes d'immigration, de citoyenneté ou de passeport, **IRCC doit être prudent pour ne pas reproduire et « ancrer » les préjugés historiques** sous le couvert de la neutralité technologique. Il ne suffit pas de comprendre nos données et notre planification sur le plan technique; nous devons assurer une collaboration interdisciplinaire pour veiller à l'équité de manière concrète.

❖ Tous les outils analytiques sont limités par leur capacité de prévoir et de décrire les résultats. Bien que les tendances générales soient utiles, elles ne peuvent pas remplacer la preuve d'actions personnelles. C'est pourquoi **IRCC doit être très prudent lorsqu'il utilise de l'information sur un groupe ou une personne**.

❖ Il est possible que l'analytique laisse supposer qu'un demandeur représente un faible risque parce que son profil personnel est similaire à celui de candidats approuvés par le passé, mais dans ce cas, **IRCC devrait gérer le risque d'une approbation automatisée (c'est-à-dire prendre en charge le risque)**. **Cependant, le calcul n'est pas le même dans le scénario inverse** : si l'analytique estime que le profil d'un demandeur est similaire à celui de candidats refusés par le passé, et recommande un refus, alors IRCC devrait interdire un refus automatisé (c'est-à-dire que le client prend en charge le risque). Comme dans le cas du modèle de décisions prises uniquement par un humain, **un refus doit être fondé sur des données probantes sur la personne en question et non seulement sur la similarité de son profil avec le profil général d'autres demandeurs** qui ont été refusés par le passé ou sur son niveau risque indiqué autrement. Par exemple, un refus automatisé fondé sur le fait que le passeport d'un demandeur de visa de résidence temporaire figure dans la base de données des documents de voyage perdus et volés d'Interpol serait justifiable, car il s'appuie sur un fait vérifiable et objectif sur le demandeur en question (c'est-à-dire qu'il ne respecte pas l'exigence d'un document de voyage valide). À l'inverse, il ne serait pas justifiable qu'un refus automatisé s'appuie sur le fait que le pays d'origine du demandeur affiche un taux élevé de fraude de passeport. Il peut s'agir d'un indicateur de risque parmi d'autres, mais ce n'est pas une raison suffisante en soi pour justifier un refus catégorique.

6. *Les agents devraient être informés, et non menés vers une conclusion.*

❖ Les agents peuvent s'appuyer sur les systèmes de soutien aux décisions en usant de leur jugement. Dans ces cas, les systèmes doivent être conçus de manière à ce qu'**ils n'entravent pas les décideurs** dans l'exercice de leur pouvoir en les orientant vers des résultats particuliers.

❖ **Les agents ont besoin d'une formation et d'un contexte appropriés pour comprendre ce que le système de soutien aux décisions leur indique précisément**. Par exemple, les étiquettes ont une importance : lorsque le système trie automatiquement les demandes des clients en groupes, il est important de savoir si ces groupes portent l'étiquette « vert/jaune/rouge » ou « simple/moyen/complexe ». La première étiquette peut sembler représenter des instructions à l'intention des agents (vert=approuver, rouge=refuser) et la dernière, une simple indication du temps et des efforts requis que devra consacrer l'agent pour prendre une décision.

❖ Un système qui **n'offre aucune occasion réelle aux agents de réfléchir** est de fait un système de décision automatisé, même si ce sont les agents qui cliquent sur le dernier bouton.

❖ Il est important d'établir un **degré élevé de cohérence entre la manière dont les agents entrent des données et la manière dont les résultats de l'analyse de données leur sont présentés**. Les concepteurs de système ont besoin que les notes de dossier des agents soient cohérentes et claires et les agents ont besoin de comprendre ce que le système veut dire lorsqu'il transmet de l'information pour appuyer leur prise de décisions.

7. *Les humaines et les systèmes algorithmiques jouent des rôles complémentaires. IRCC devrait chercher continuellement à optimiser ces rôles et à trouver l'équilibre adéquat pour tirer le meilleur de chacun.*

❖ **Les ordinateurs devraient miser sur leurs forces**, soit d'analyser avec fiabilité de grands volumes de données et d'examiner des millions de possibilités sans se fatiguer, et **les humains devraient miser sur les leurs**, soit leur

7

intuition, leur créativité, leur empathie, leurs compétences sociales et leur capacité de façonner une stratégie globale. Les humains sont meilleurs pour créer les questions et les machines sont meilleures pour trouver les réponses. **Les systèmes algorithmiques devraient permettre aux humains de se concentrer sur des aspects qui requièrent leur expertise et leur jugement.**

❖ **Les systèmes qui prennent eux-mêmes les décisions administratives**, au lieu d'appuyer les humains à prendre une décision, **conviennent pour les décisions concernant des déterminations simples et factuelles**, notamment le fait que le client possède un dossier criminel ou est déjà venu au Canada. L'automatisation devrait être centrée d'abord et avant tout sur des éléments routiniers, où il est facile de décrire les résultats souhaités et où il n'est pas nécessaire que l'humain exerce son jugement.

❖ Au moment de déterminer si une tâche ou une étape dans le processus de prise décisions doit passer de l'humain à la machine, les concepteurs de programmes devraient se poser la question suivante : **« dans cette situation, des esprits raisonnables pourraient-ils aboutir à une conclusion différente? »** Si oui, alors l'automatisation n'est pas recommandée. À l'inverse, s'il semble que pratiquement tous les agents tireraient la même conclusion selon les faits dont ils disposent, l'automatisation est recommandée.

❖ **Non seulement la présence d'un humain dans le processus assure une protection contre les risques à l'équité procédurale, mais elle produit aussi de meilleurs résultats.** Pour les évaluations simples (par exemple la mise en correspondance d'une personne et de sa photo), le rendement des algorithmes fondés sur les données est souvent supérieur à celui de l'humain. Quand toutefois les machines et les humains unissent leurs forces, les résultats sont encore meilleurs. L'une des façons d'améliorer le rendement est de demander à des experts d'appliquer leur jugement dans les extrants de modèles mathématiques; une autre consiste à inverser la séquence et demander aux experts et aux décideurs de contribuer à la conception du modèle dès le départ.

❖ **IRCC devrait régulièrement examiner et ajuster ses systèmes automatisés** pour s'assurer qu'ils reflètent la réalité sur le terrain et fonctionnent conformément aux priorités stratégiques du jour. Des boucles de rétroactions sont essentielles à une conception responsable et au contrôle de la qualité. Par exemple, la rétroaction des agents de première ligne doit se rendre aux concepteurs de systèmes pour qu'IRCC puisse continuellement réévaluer si l'on peut se fier aux conclusions d'un système. De façon similaire, parce que les systèmes de décisions automatisées ne peuvent pas rétablir les priorités stratégiques eux-mêmes, les concepteurs de systèmes devront les ajuster continuellement pour tenir compte des changements de direction.

❖ Même si IRCC finit par s'appuyer largement sur le soutien aux décisions automatisé pour répondre à la demande, **il est essentiel que des agents humains continuent d'examiner une portion des demandes manuellement.** Ainsi, les capacités des agents ne s'atrophieront pas et les connaissances d'IRCC sur les conditions locales d'un pays et les tendances en matière de fraude resteront à jour.

❖ **Les experts techniques (et leur gestionnaire) sont en position de confiance et ont la responsabilité particulière d'évaluer en profondeur des systèmes informatiques et leur incidence potentielle.** Comme toujours, le Ministère s'attend à des évaluations objectives et des conseils francs. Avec la montée des systèmes d'apprentissage machine, il faut faire extrêmement attention de cerner, d'expliquer et d'atténuer les risques potentiels, surtout si ces risques ne sont apparents que pour les personnes comprenant les aspects techniques des systèmes et des modèles en jeu. Les gestionnaires doivent créer des conditions dans lesquelles les experts peuvent consulter leurs pairs, réévaluer les risques à mesure qu'évoluent les systèmes et signaler les dangers sans peur de représailles.

8

❖ IRCC devrait déployer tous les efforts pour **donner à ses employés les compétences, les outils et le soutien dont ils ont besoin** pour innover. Par soutien, on entend le temps accordé aux employés pour qu'ils apprennent et l'espace pour qu'ils expérimentent de façon responsable.

8. *En plus de respecter le cadre de protection de la vie privée du Canada, IRCC devrait continuellement adopter des pratiques exemplaires en matière de protection de la vie privée dans un domaine qui évolue rapidement.*

❖ En raison de la tension potentielle entre les technologies fondées sur les données et les mécanismes de protection de la vie privée, les scientifiques de données, les concepteurs de programmes et les spécialistes en TI devraient parfaitement comprendre les enjeux touchant la protection de la vie privée, ainsi que les <u>droits et responsabilités associés à la collecte, l'utilisation, la divulgation et la rétention de renseignements personnels</u>, ainsi que **les normes et pratiques exemplaires émergentes relatives à l'utilisation d'autres renseignements sur les personnes.** IRCC devrait envisager d'offrir de la formation obligatoire à cet effet, car il peut être insuffisant de laisser le personnel déterminer eux-mêmes leurs besoins de formation sur la protection de la vie privée dans cette ère de données massives.

❖ IRCC transmet des faits sur des clients individuels à certains partenaires canadiens et internationaux et il aimerait aussi communiquer de l'information sur ses systèmes automatisés, y compris, dans certains cas, les algorithmes. Cependant, en règle générale, IRCC ne doit pas transmettre de prévisions générées par un système (comme les notes, signalements ou recommandations liés aux risques) à propos de clients individuels, à moins : **A) que la communication soit exigée par la loi (p. ex. conformément à la *Loi sur la communication d'information ayant trait à la sécurité du Canada*), B) que nous fournissions le contexte intégral, y compris une base pour que les partenaires comprennent et évaluent le bien-fondé de la conclusion statistique, et C) qu'il y ait une façon de reprendre, corriger ou modifier l'information si les circonstances changent.** Il s'agit d'un moyen de protection dans le cas d'un préjugé algorithmique accidentel. Si un tel préjugé désavantage une personne, alors un partage élargi entre les partenaires peut produire un effet de domino, multipliant la discrimination vécue par cette personne. Il faut examiner attentivement toute exception aux règles de non-partage et obtenir l'approbation au niveau du sous-ministre.

❖ Le Ministère devrait être **proactif et se préparer à une violation de données** relativement à l'un ou à plusieurs de ses systèmes algorithmiques. Cela signifie qu'il doit avoir la bonne équipe et les procédures normales d'exploitation adéquates en place.

TRANSPARENCE ET EXPLICABILITÉ

9. *IRCC doit soumettre tous les systèmes à une surveillance continue pour s'assurer qu'ils sont techniquement adéquats, conformes aux exigences juridiques et en matière de politiques, justes et fonctionnels comme prévu.*

❖ Pour assurer la légitimité des systèmes, il est important d'encourager le **contrôle externe** de leur conception. Des experts externes peuvent nous aider à veiller à la santé technique de nos systèmes et à l'absence de préjugés involontaires. Lorsque possible, les nouveaux systèmes algorithmiques devraient être accessibles aux experts externes, comme un comité consultatif, à des fins de vérification, d'essai et d'examen. Si les essais ne sont menés qu'à l'interne, alors il faudrait documenter les méthodes et les hypothèses mises de l'avant lors des essais, ainsi que les résultats qui en découlent, puis les rendre accessibles au comité consultatif.

9

10. IRCC doit toujours expliquer clairement ses décisions relatives aux demandes de clients.

❖ Peu importe le degré auquel IRCC s'appuie sur ses systèmes algorithmiques pour soutenir sa prise de décisions, les demandeurs jouissent du même droit qu'aujourd'hui de comprendre le fondement des décisions prises quant à leur demande. Ces explications sont essentielles pour le demandeur qui veut profiter de mécanismes de recours et pour assurer la responsabilisation du gouvernement en général (les députés et les journalistes aussi s'attendent à une explication détaillée des décisions prises par IRCC). Il peut être difficile sur le plan technique d'expliquer le fonctionnement des systèmes algorithmiques, mais **le public ne devrait pas avoir besoin de comprendre ce dernier pour qu'on lui explique les décisions automatisées**. Si une explication procure peu de renseignements utiles, alors elle offre peu de valeur.

❖ Une des façons de rendre des explications compréhensibles pour les humains est de limiter le rendement d'un système aux fonctions qui sont facilement explicables. Toutefois, cela peut signifier qu'il faudra gaspiller le potentiel des systèmes d'IA qui peuvent comprendre les données complexes de manières dont les humaines ne peuvent pas. **Au lieu d'imposer des limites strictes sur ce que les outils peuvent *faire*, IRCC devrait tenter de trouver de nouvelles et meilleures façons d'*expliquer* ce qu'ils peuvent faire.**

❖ Étant donné l'interdiction des systèmes « boîtes noires » pour la prise de décisions administratives complètes, l'utilisation des systèmes de soutien aux décisions automatisé devrait rendre les décisions **plus facilement explicables, pas moins**. Par exemple, on peut utiliser des pistes de vérification tenant compte de tous les facteurs et étapes dans le processus de prise de décisions, au besoin, pour générer et consigner systématiquement les raisons d'une approbation ou d'un refus.

❖ Une **piste de vérification** est vitale pour respecter le droit d'un demandeur de comprendre le fondement de la décision prise concernant sa demande, qu'elle soit positive ou négative, et peut être utile pour satisfaire un tribunal dans le cas d'une contestation juridique. Tout nouvel outil incapable de **générer automatiquement une piste de vérification sur le cheminement de la prise de décisions administratives** n'est pas prêt à la mise en œuvre.

❖ Tout comme dans le cas des explications des décisions administratives individuelles, **IRCC doit trouver des façons d'être manifestement transparent sur le fonctionnement des systèmes dans leur intégralité**. Bien que les stratégies pour expliquer le travail du Ministère soient différentes selon le public cible, il faut au moins :
   — Que les clients comprennent comment leur demande sera traitée et leurs renseignements personnels utilisés
   — Les Canadiens doivent comprendre comment IRCC dépense l'argent des contribuables pour faciliter l'entrée de familles, d'amis et de visiteurs légitimes tout en protégeant le public contre les menaces à la santé et à la sécurité.
   — Les intervenants, les médias et les experts techniques ont besoin d'une base pour faire une critique éclairée de l'approche du Canada.

11. IRCC doit équilibrer sa transparence et son besoin de protéger la sécurité des Canadiens.

❖ Il est dans l'intérêt du Ministère de ne pas divulguer trop d'information sur ses systèmes de décisions pour des raisons d'intégrité des programmes. En divulguant trop d'information, il pourrait faciliter la tâche aux personnes voulant manipuler la prise de décision ou « contourner le système ». Il doit fournir des explications sur le fonctionnement des systèmes et sur les décisions qu'ils aident le Ministère **suffisamment détaillées pour**

10

indiquer de manière générale à un client comment une décision est prise, mais pas trop détaillées au point de compromettre l'intégrité des programmes et, par extension, la sécurité des Canadiens.

❖ Étant donné que toutes les règles de prise de décisions pourraient ultimement se faire connaître du public, **IRCC devrait viser à concevoir des systèmes, dans la mesure du possible, en s'appuyant sur des points de données et des facteurs qui sont difficiles à falsifier ou à inventer**. Par exemple, des dossiers bancaires sont plus difficiles à falsifier qu'un motif de voyage, que le demandeur n'a qu'à déclarer. Du mieux que nous le pouvons, nous devrions nous appuyer sur les faits objectifs pour atténuer le risque que le demandeur mente pour contourner les règles. Le degré d'importance accordé aux différents éléments de données dans les algorithmes d'IRCC devrait tenir compte du taux de tentative de fraude dont ils ont été la cible par le passé ou de la probabilité qu'ils soient ciblés à nouveau à l'avenir.

12. *Les clients continueront d'accéder aux mêmes mécanismes de recours et le fait qu'IRCC utilise des systèmes automatisés ne devrait pas réduire la capacité d'une personne à miser sur ces recours.*

❖ Lorsqu'un client conteste une décision qui a été prise en tout ou en partie par un système automatisé, et qu'IRCC a le pouvoir et accepte d'examiner la décision, **alors ce sera un humain qui effectuera cet examen.**

❖ Toutefois, **IRCC ne devrait pas offrir proactivement aux clients, au début du processus de demande, l'option qu'un agent examine leur dossier et prenne une décision à cet égard au lieu du système automatisé**. L'objectif d'IRCC est de mettre en œuvre des systèmes responsables, efficaces et efficients. En permettant aux clients de choisir d'éviter les systèmes, nous compromettrions la capacité du Ministère d'assurer un traitement rapide et cohérent, et cela laisserait entendre que les systèmes automatisés offrent un service inférieur à celui fourni par les agents humains.

11

3. LE MANUEL DU RESPONSABLE DE L'AUTOMATISATION

A. Explorer le soutien aux décisions automatisé à titre de solution

La présente section répondra à des questions comme celles-ci : *Le problème que vous tentez de résoudre se prête-t-il à une solution fondée sur les données? Avez-vous ce dont vous avez besoin pour réussir? Les avantages surpassent-ils les risques?*

Planifiez-vous l'automatisation de processus simples ou l'intervention que vous proposez contribuera-t-elle, de quelque manière que ce soit, à la prise de décisions concernant l'approbation des demandes des clients?

Il peut sembler que certaines étapes d'un flux de travail soient retirées de la prise de décisions définitives, mais en réalité, elles peuvent être partiellement déterminantes dans l'approbation ou le refus. Par exemple, si un système automatisé répartit des demandes reçues dans trois compartiments – vert pour les demandes simples, jaune pour les demandes moyennes et rouge pour les demandes complexes – alors les agents peuvent percevoir, consciemment ou inconsciemment, les étiquettes verte, jaune ou rouge comme des recommandations sur le degré de risque des demandes, et donc, sur celles qu'ils devraient approuver.

Les spécialistes des politiques et programmes cherchant à automatiser toute partie d'un processus de décision devraient se demander si l'automatisation qu'ils proposent est purement centrée sur le processus, par exemple en informatisant des tâches répétitives de moindre importance qui étaient auparavant accomplies manuellement, ou si elle peut influer sur quelque chose de plus important.

SI L'AUTOMATISATION EST PUREMENT CENTRÉE SUR UN PROCESSUS　　⟶　　LES SECTIONS SUIVANTES DE LA PRÉSENTE POLITIQUE NE S'APPLIQUENT PAS (Ajoutez votre projet à **cette liste** de systèmes automatisés qu'utilise IRCC)

SI ELLE EST SUSCEPTIBLE DE CONTRIBUER À LA PRISE DE DÉCISIONS　　⟶　　CONTINUEZ À LIRE

Proposez-vous que le système formule des recommandations directement aux agents sur l'approbation ou le refus d'une demande, ou que le système aboutisse à une décision définitive sans intervention humaine?

NON　　⟶　　CONTINUEZ À LIRE
LA PRÉSENTE POLITIQUE S'APPLIQUE AUX SYSTÈMES QUI APPUIENT LA PRISE DE DÉCISIONS D'AUTRES MANIÈRES

OUI　　⟶　　CONTINUEZ À LIRE
VOUS SEREZ AUSSI ASSUJETTI AUX EXIGENCES DE LA *DIRECTIVE SUR LA PRISE DE DÉCISIONS AUTOMATISÉE* DU SCT.

Planifiez-vous un système qui utilise l'apprentissage machine pour générer et ajuster de manière autonome des règles opérationnelles sans intervention humaine?

NON　　⟶　　CONTINUEZ À LIRE
LA PRÉSENTE POLITIQUE S'APPLIQUE AUSSI AUX SYSTÈMES FONDÉS SUR DES RÈGLES FIXES

12

| OUI | → | CONTINUEZ À LIRE AVEC ATTENTION
VOTRE PROJET DEVRA NÉCESSITER UNE ATTENTION SPÉCIALE ET DÉCLENCHER LES EXIGENCES LES PLUS STRICTES DE LA *DIRECTIVE SUR LA PRISE DE DÉCISIONS AUTOMATISÉE* DU SCT. |

Si votre objectif est la prise de décisions automatisées, avez-vous réfléchi à savoir si l'automatisation de processus simple ou de nouvelles approches à la répartition de la charge de travail peuvent atteindre des gains d'efficacité similaire? Avez-vous épuisé toutes les options simples avant d'envisager le domaine complexe et litigieux du soutien aux décisions automatisé?

Aurez-vous besoin d'un nouveau pouvoir en matière de politiques ou juridique?

Le Ministère possède un vaste pouvoir d'utiliser des outils électroniques pour réaliser son mandat (voir l'article 4.1 de la *Loi sur l'immigration et la protection des réfugiés* [LIPR] et les modifications à l'article 2.2 du *Décret sur les passeports canadiens*), mais vous pourriez avoir besoin de nouveaux pouvoirs si vous prévoyez une automatisation entraînant des changements importants aux exigences de programme (par exemple les critères de recevabilité), aux résultats de programme (par exemple, quels clients sont approuvés) à la collecte, l'utilisation ou la communication de renseignements personnels, ou au rôle des agents dans le processus de prise de décisions.

Il se peut que vous deviez travailler avec les responsables des politiques pour obtenir éventuellement de nouveaux pouvoirs au moyen d'un mémoire au Cabinet ou d'autres instruments.

Avez-vous réfléchi à propos des personnes qui doivent participer à la discussion?

Vous devrez former un groupe pluridisciplinaire intersectoriel, qui comprendra probablement :

La DGOPI et les réseaux de traitement	Possèdent des connaissances pratiques du processus de prise de décisions, du rôle des agents, des coûts, des gains d'efficacité potentiels, de l'évolutivité, etc.
Planification et rendement des opérations *Équipe de gouvernance des données et laboratoire d'analytique avancée*	Aident à l'évaluation de la qualité et au caractère adéquat des données et de l'expertise en science appliquée.
Politique stratégique et planification *Équipe de politiques numériques*	Donne des conseils sur les pouvoirs juridiques et en matière de politiques, la conception responsable, les liens avec d'autres initiatives au Ministère, la Directive sur la prise de décisions automatisée du SCT, etc.
Une direction générale sur les politiques de programmes, selon le dossier	Aborde les changements possibles aux politiques ou programmes pour compléter votre initiative.
Services juridiques	Donnent des conseils sur les exigences des programmes, la protection de la vie privée, la *Charte*, le droit administratif, la propriété intellectuelle et les risques de litige.
Direction générale de la stratégie numérique (Secteur de la transformation et des solutions numériques)	Aborde la manière dont votre initiative cadre avec la stratégie générale d'IRCC en matière de technologie.
Direction générale des projets (Secteur de la transformation et des solutions numériques)	Est responsable de la gestion et de l'établissement de l'ordre de priorité quant à l'intégration des projets de GI-TI.
AIPRP et responsabilisation *Division de l'AIPRP*	Donne des conseils sur la nécessité d'effectuer ou de réviser une évaluation des facteurs relatifs à la vie privée (EFVP), un

13

	énoncé de confidentialité ou un fichier de renseignements personnels (FRP).
Communications	Déterminent si des communications au public sont nécessaires concernant votre projet.
Expérience client ** Pour des propositions qui pourraient toucher directement les interactions avec les clients*	Donne des conseils sur la conception et la mise à l'essai de l'expérience utilisateur

Vous êtes-vous demandé si vous avez besoin de la participation de l'ASFC ou d'un autre ministère fédéral? Même si la présente politique cible les activités d'IRCC, votre projet pourrait avoir une incidence sur ses partenaires (qui peuvent aussi avoir en place des politiques et approches en matière d'analytique, d'IA et d'automatisation).

IRCC possède-t-il des données d'une portée et d'une qualité suffisantes pour que votre proposition soit viable?

— Y a-t-il suffisamment de données structurées de grande qualité pour appuyer des conclusions fiables? Quelles sont les limites des données? Sont-elles exactes, à jour et véritablement représentatives, et contiennent-elles des préjugés historiques qui pourraient être maintenus? *Vous devrez faire évaluer la qualité de vos ensembles de données avec l'aide du leadership de la fonction de gouvernance des données ministérielle et de la collaboration entre les opérations et les TI. Cette évaluation devrait être approuvée par un groupe de gestion de la gouvernance des données, comme le comité de direction des données.*

— Combien de vos points de données sont des intermédiaires pour les critères ou résultats réels que vous visez? Tous? *Vous aurez besoin d'indicateurs éprouvés indiquant que les demandeurs sont conformes aux objectifs ou exigences du programme.*

— Pouvez-vous utiliser les données? Sont-elles invasives ou posent-elles des risques élevés au droit à la protection de la vie privée des personnes? Selon quel fondement utiliseriez-vous les données dans votre analyse dans une application du monde réel? Avez-vous besoin d'un accès à de l'information qu'IRCC recueille déjà (il faudrait examiner l'utilisation interprogrammes de l'information déjà recueillie pour un programme), ou cherchez-vous à tirer des données externes, comme celle que détiennent les partenaires fédéraux, provinciaux ou internationaux? *Vous devrez obtenir des conseils auprès de la division d'AIPRP et possiblement des Services juridiques pour déterminer si la collecte, l'utilisation ou la divulgation des données que vous proposez sont conformes aux cadres de protection de la vie privée régissant le Ministère ou si vous devrez obtenir de nouveaux pouvoirs.*

Votre projet requiert-il une attention particulière?

❑ S'inscrit-il dans un domaine sujet à l'examen minutieux du public (en raison des préoccupations touchant la vie privée, par exemple) ou à des litiges fréquents?

❑ Les clients dans ce secteur d'activité sont-ils particulièrement vulnérables?

❑ Les enjeux des décisions sont-ils très élevés?

❑ Envisagez-vous des refus automatisés?

❑ Cela aura-t-il une incidence importante sur les employés d'IRCC, que ce soit sur le plan de leur nombre ou de leur rôle?

❑ Envisagez-vous d'utiliser un algorithme « boîte noire »?

Comment savoir si vous avez une « boîte noire »?

❑ L'algorithme prévisionnel est-il protégé contre la divulgation (par exemple un secret commercial)?

❑ Le processus n'est-il pas décrit avec précision?

❑ Le processus d'analyse ou de justification du système est-il inconnu? Son comportement est-il difficile à interpréter ou expliquer?

14

❑ Envisagez-vous de recueillir ou d'utiliser des données non traditionnelles ou potentiellement sensibles, comme des données associées aux médias sociaux des clients?

Si vous avez répondu par l'affirmative à une ou plusieurs questions, votre projet nécessitera des consultations et une surveillance accrues.

Quelles incidences, négatives et positives, l'automatisation proposée aura-t-elle sur les clients? *Par exemple :*

❑ Permettra-t-elle un traitement accéléré?
❑ Exigera-t-elle des clients qu'ils confient à IRCC davantage de renseignements personnels ou touchera-t-elle autrement leur vie privée?

Quelles incidences, négatives et positives, l'automatisation proposée aura-t-elle sur le Ministère? *Par exemple :*

❑ Lui fera-t-elle faire des économies (à court, à moyen ou à long terme)?
❑ Améliorera-t-elle la cohérence du processus?
❑ Améliorera-t-elle l'intégrité de ses programmes?
❑ Augmentera-t-elle le risque de litige?
❑ Changera-t-elle les compétences essentielles, les descriptions d'emploi ou les classifications des décideurs au sein d'IRCC? Changera-t-elle la taille et la composition du personnel requis pour exécuter un programme?

Avez-vous réfléchi à la réaction des membres du personnel de première ligne?

Il serait utile de les consulter et de consulter les ressources humaines. Même les propositions qui ne prévoient pas de grands changements au rôle des agents peuvent provoquer chez eux de l'incertitude et du stress. En veillant dès le début à communiquer les objectifs du projet, vous rassurerez le personnel, gagnerez du temps et des efforts plus tard.

Avez-vous réfléchi à la réaction des intervenants externes?

Il ne sera pas pratique de consulter des intervenants externes à propos du développement de chaque algorithme ou système automatisé qu'utilise IRCC, mais il est important de tenir compte des perspectives de ceux qui seront touchés par les aspects automatisés du processus de prise de décision, particulièrement lorsque de nouveaux algorithmes ou systèmes sont intégrés à un tout nouveau programme ou à un changement substantiel de l'approche de prestation.

Les intervenants traditionnels d'IRCC sont toujours importants, mais vous devrez tenir compte d'un plus grand éventail d'intervenants, comme ceux qui travaillent dans le domaine de l'éthique des technologies ou de l'IA.

Il est recommandé que vous consigniez les perspectives des intervenants comme vous le feriez dans le cadre du développement d'un changement important à une politique ou une loi.

Savez-vous si vous devrez entreprendre une *évaluation des facteurs relatifs à la vie privée*, ou réviser une évaluation existante?
La division de l'AIPRP peut vous aider à déterminer ce qui convient.

Avez-vous tiré profit de la formation offerte sur les technologies fondées sur les données et sur leurs effets?

Avez-vous préparé plus d'une option?

Il est peu probable qu'il n'y ait qu'une façon d'atteindre votre objectif. En explorant d'autres options, en établissant des modèles connexes et en les mettant à l'essai, vous pourriez confirmer votre hypothèse initiale et renforcer votre analyse de rentabilité ou découvrir une meilleure approche. N'oubliez pas que le cadre de protection de la vie s'applique même pendant les étapes de la création du modèle de vos activités.

Le modèle que vous avez établi a-t-il généré de nouvelles connaissances sur la conception possible de votre programme?

Par exemple, l'expression du système d'intérêt facilitée par le programme Entrée express a soulevé des occasions de moderniser le programme économique fédéral dans son ensemble. De plus, l'analytique de données sous-jacente au modèle prévisionnel de visa de résidence temporaire lié à la Chine a renforcé la justification du programme CAN+ (soit l'historique des voyages).

— La solution que vous proposez remplace-t-elle efficacement les exigences de programmes établies dans les lois ou les règlements par des règles opérationnelles générées par des algorithmes?

Avez-vous bien prévu à votre budget le développement de systèmes d'itération et d'expérimentation d'analytique de données dans votre mémoire au Cabinet ou présentation au Conseil du Trésor (si nécessaire)?

Il faut consacrer du temps et de l'argent pour révéler de nouvelles connaissances opérationnelles au moyen de l'analytique avancée. Le processus pour créer un nouveau système de soutien aux décisions automatisé peut comporter des essais et des erreurs et en fixant un budget à cet égard, vous aurez la flexibilité de faire les itérations requises avant de réaliser le produit final visé. Une fois construit, un système automatisé a besoin de ressources pour sa maintenance et son renouvellement périodique.

B. Concevoir votre système

La présente section répondra à des questions comme celles-ci : *Comment assurons-nous l'équité dans notre conception? Comment protégeons-nous l'impartialité des décideurs? Comment répondons-nous aux attentes des clients en matière de transparence et de protection des renseignements personnels?*

Quelles étapes suivrez-vous pour minimiser les préjugés involontaires?

Des algorithmes mal conçus peuvent introduire des préjugés, mais dans la plupart des cas, les préjugés proviennent des données utilisées pour l'analytique ou la formation des modèles d'apprentissage machine. Les concepteurs de programmes doivent être conscients des faiblesses dans les données au cœur de leur projet. Parmi les techniques visant à atténuer les préjugés dans les données, notons la consultation d'experts pour veiller à ce que les données soient situées dans le bon contexte (par exemple, IRCC a-t-il changé certaines définitions opérationnelles au fil du temps?) et la formation contradictoire exigée pour les modèles d'apprentissage machine qui pourrait corriger toute injustice perçue. La mise à l'essai rigoureuse de l'assurance de la qualité peut aussi aider à détecter les tendances qui suggèrent les préjugés et les corriger immédiatement.

Peu importe les détails d'un projet donné, il est toujours conseillé de réunir différentes perspectives et identités au sein de votre équipe de développement. Il existe de nombreux exemples de projets d'IA qui ont connu des ratés en raison du manque de diversité dans le groupe d'experts. Si vous ne tenez pas compte des circonstances uniques des différents groupes de gens concernés, cela risque de nuire aux populations que votre système est pourtant conçu pour aider. Pour vous protéger contre des préjugés involontaires, vous devez avoir un point de vue d'ACS+ pour votre analyse de données et les premiers travaux de conception. *Vous devriez vous efforcer, le plus possible, d'avoir au sein de votre équipe des membres qui ont différentes vues du monde, peu importe le sexe, le genre, l'âge, l'ethnie, la langue, la ruralité, la capacité ou d'autres facteurs d'identité.*

Étant donné le besoin de transparence, particulièrement dans le contexte des contestations juridiques, proposez-vous des règles opérationnelles qui sont justifiables et rendent la tâche difficile aux gens qui veulent les manipuler?

— Les modèles et les règles doivent aborder les exigences de programmes de sorte que les décisions soient fondées sur ces exigences et sur de l'information pertinente et fiable.

— Concevez-vous votre système, dans la mesure du possible, en fonction de facteurs ou de points de données qui sont difficiles à falsifier ou à inventer? Il suffit aux clients d'indiquer la plupart des renseignements sur leur demande, sans qu'IRCC puisse l'authentifier. Toutefois, il est possible de vérifier d'autres points de données, comme les dossiers bancaires ou des visas ayant déjà été accordés. En privilégiant ces points de données, on contribue à l'atténuation du risque que le demandeur qui connaît le fonctionnement des systèmes d'IRCC mente pour contourner les règles.

— Dans la pondération des variables dans un algorithme, tenez-vous compte du degré auquel chaque variable a pu être la cible de fraude par le passé ou sera potentiellement ciblée à l'avenir?

> **Le saviez-vous?**
> Vous pouvez consulter la Direction de consultation du gouvernement du Commissariat à la protection de la vie privée (CPVP) du Canada pour obtenir des conseils sur les questions de vie privée, même si vous ne menez pas une évaluation des facteurs relatifs à la vie privée. Pour ce faire, commencez par contacter la Direction de l'AIPRP et responsabilisation à IRCC.ATIPInternal-InterneAIPRP.IRCC@cic.gc.ca . Le responsable de la protection des renseignements personnels d'IRCC est le point de contact central pour traiter avec le CPVP.

Comment votre système teindra-t-il compte des <u>principes de protection de la vie privée dès la conception</u>?

— Tenez-vous compte de la protection de la vie privée à toutes les étapes du développement du programme?

— Avez-vous été proactif en intégrant la protection de la vie privée dès la conception de l'architecture de votre système?

Dans votre approche à l'utilisation des renseignements personnels à des fins légitimes, sans violation des droits des personnes ou des groupes, vous devriez :

— Comprendre la provenance des données

— Assurer leur exactitude

— Réduire les données marginales, redondantes ou inutiles

— Prendre des précautions pour éviter que des données anonymisées (ou pseudoanonymisées) soient de nouveau personnalisées et prévenir la collecte de données non autorisée

— Protéger les données contre l'accès non autorisé et la divulgation accidentelle

Les politiques de données et les procédures de protection de la vie privée transparentes devraient permettre aux personnes ce qui suit :

— Comprendre quelles données sont recueillies, comment elles sont utilisées et divulguées et pendant combien de temps elles sont conservées

— Donner leur consentement éclairé quant à la collecte de données automatique et pour tout partage de données après coup

— Examiner et obtenir leurs renseignements personnels et corriger toute erreur

Comment concevrez-vous l'interaction humain-machine?

— Votre système remplacera-t-il l'humain décideur ou l'aidera-t-il simplement?

— S'il s'agit d'un outil de soutien, quelles étapes suivez-vous pour protéger l'impartialité des décideurs? Par exemple, vous devez veiller à ce qu'ils comprennent bien ce que leur dit le système de soutien aux décisions. Les extrants du système devront être présentés comme des conseils spécialisés et non comme des décisions déjà prises (où la machine est de fait le décideur).

— Comment votre conception maximise-t-elle la valeur des humains? Par exemple, un agent pourrait-il se voir confier un dossier complexe?

— Dans quelles circonstances un agent pourrait-il déroger à une décision prise par le système?

Votre système utilisera-t-il un logiciel de source ouverte?

La <u>Directive sur la gestion des technologies de l'information du SCT</u> exige que les ministères utilisent des logiciels et normes de sources ouvertes par défaut, sauf si l'option de source ouverte n'est pas accessible ou ne répond pas aux besoins de l'utilisateur.

S'ils n'utilisent pas des sources ouvertes, les ministères doivent privilégier les options commerciales non contraintes aux plateformes par rapport aux produits commerciaux exclusifs pour éviter la dépendance à une technologie particulière et pour que soient possibles la substitution et l'interopérabilité.

Si une application personnalisée est développée, par défaut, tout code source écrit par le gouvernement doit être publié en format ouvert sur les sites Web du gouvernement du Canada et les services désignés par le SCT.

18

Comment votre système sera-t-il conçu pour assurer l'équité procédurale?

L'équité procédurale exige que les demandeurs aient une *occasion suffisante de présenter leur dossier à un décideur impartial*. Les exigences peuvent varier en pratique, mais il peut s'agir de ce qui suit pour les demandeurs :

— Avoir la possibilité de faire des présentations
— Se faire divulguer les preuves négatives et avoir la chance de se faire entendre à cet égard
— Se faire divulguer les raisons d'une décision négative

IRCC utilise une échelle variable; plus les risques associés à une décision sont élevés, plus grande est la chance que l'on accorde aux clients pour qu'ils expliquent leurs préoccupations avant que l'on rende une décision.

Afin de protéger ou d'améliorer l'équité procédurale, vous devrez réfléchir attentivement à la façon dont l'ajout d'un soutien aux décisions automatisé changera le traitement des demandes et la prise de décisions connexes. Par exemple :

❑ Si vous envisagez un modèle prévisionnel, comment informera-t-il les agents sans entraver l'exercice de leur pouvoir discrétionnaire?
❑ Tous les facteurs pertinents sont-ils pris en compte?
❑ Votre système tirera-t-il de l'information ailleurs que celle fournie par le demandeur? Les demandeurs ont le droit de savoir quelles autres preuves externes ont été utilisées lors de la prise de décisions, en partie pour qu'ils aient la possibilité de corriger toute information erronée.
❑ Quelles raisons seront fournies lorsqu'une décision est prise?

Assurer des explications claires et utiles

Grâce à l'équité procédurale, les demandeurs ont le droit de comprendre le fondement de la décision prise à l'égard de leur demande. L'explication qu'ils reçoivent est essentielle à leur capacité de profiter de mécanismes de recours et à la responsabilisation du gouvernement en général.

IRCC n'a pas encore mis en œuvre les refus automatisés des demandes de clients. S'il en venait à le faire, les concepteurs de programmes devront porter attention à ce qui suit :

— Pour les décisions administratives prises par des systèmes de soutien aux décisions automatisé, ou avec l'aide de ceux-ci, il faut au moins pouvoir les expliquer aussi aisément que les décisions traditionnelles prises seulement par des humains. C'est pourquoi les algorithmes « boîtes noires », comme des réseaux neuronaux avancés, ne sont pas recommandés comme moyen principal de déterminer si une demande est approuvée ou non.

— Les explications ne doivent pas compromettre l'intégrité du programme. Heureusement, il n'est généralement pas nécessaire d'être transparent au point de divulguer le fonctionnement interne d'un système d'IA. Plutôt, on doit fournir des explications qui répondent à des questions comme celles-ci : *Quel est le fondement de la décision? Quels sont les principaux facteurs? En changeant un facteur particulier, aurait-on obtenu une décision différente? Pourquoi deux dossiers similaires obtiennent-ils des décisions différentes?*

— En règle générale, les explications sur les décisions automatisées devraient : (1) aider le client à comprendre pourquoi une décision donnée a été obtenue, (2) lui fournir les motifs pour contester la décision s'il le désire. Les concepteurs de programme devraient aussi déterminer s'il sera possible d'indiquer, de façon globale, ce que le client devrait changer pour qu'il obtienne le résultat qu'il souhaite à l'avenir. On appelle cette approche « explication contre-factuelle ».

* À l'heure actuelle, IRCC fournit des explications telles que : « Je ne suis pas convaincu que vous quitterez le Canada à la fin de votre séjour à titre de résident temporaire...selon votre emploi actuel » ou « ...selon votre historique de voyages ».

19

Qui sera le décideur officiel?

L'<u>article 4.1 de la LIPR</u> indique que les agents peuvent prendre des décisions ou faire des déterminations ou des examens au moyen de systèmes automatisés rendus accessibles par le ministre (des pouvoirs similaires sont visés pour le *Décret sur les passeports canadiens*, et ils peuvent être requis pour la *Loi sur la citoyenneté*). Cela ne signifie pas que les systèmes eux-mêmes deviennent les décideurs officiels. Les humains conçoivent des architectures de système, des règles opérationnelles, des seuils de fiabilité et d'autres mécanismes similaires et ce sont donc eux qui sont ultimement responsables de chaque décision que prend le Ministère. Pour sa part, le ministre est responsable devant le Parlement de toutes les activités de son ministère.

> Pour le projet d'analytique avancée relatif aux visas de résidence temporaire de la Chine, IRCC a désigné le directeur général de la Direction générale de l'orientation du programme d'immigration comme décideur officiel pour toutes les décisions concernant la recevabilité prises sans l'examen d'un agent.
>
> En règle générale, si on utilise un modèle prévisionnel ou un système automatisé pour appuyer la prise de décisions administratives, il est logique de démontrer sa responsabilisation en nommant un décideur officiel qui approuvera les règles opérationnelles qu'applique le système.
>
> Les règles devraient être décrites clairement pour qu'il n'y ait aucun doute que le décideur comprend entièrement des règles et, par extension, le fondement des décisions sous-jacentes.

— Qui approuvera les règles opérationnelles utilisées dans votre système de soutien aux décisions?

— S'agit-il du décideur officiel?

— Si un système automatisé détermine certains éléments d'une décision administrative, mais pas tous, la personne qui rend la décision définitive sera-t-elle considérée comme le décideur officiel pour tout le processus décisionnel? Ou plutôt, certains sous-éléments de la décision seront-ils considérés comme des décisions en soi ayant leur propre décideur officiel (par exemple, la recevabilité par rapport à l'admissibilité)?

Comment votre système conservera-t-il une piste de vérification pour chaque décision?

Pour se préparer à des contestations juridiques potentielles, IRCC doit absolument établir une chaîne de preuves démontrant comment les systèmes dans leur ensemble fonctionnent, ainsi qu'une piste concernant chaque décision individuelle ayant été touchée par ce système.

❑ Le système automatisé a-t-il la capacité de générer automatiquement une piste de vérification exhaustive démontrant le cheminement de la prise de décision administrative ou conserverez-vous un document électronique dans le dossier du demandeur?

> Pour l'AVE et les prolongations du permis d'études et de la fiche de visiteur, l'onglet d'automatisation dans le Système mondial de gestion des cas (SMGC) permet aux utilisateurs de voir clairement toutes les sous-activités.

❑ Tous les points de décision clés peuvent-ils être repérés dans la piste de vérification?

❑ Tous les points de décision clés s'inscrivent-ils dans la logique du système automatisé relativement aux lois, politiques ou procédures applicables?

❑ Toutes les décisions sont-elles consignées et accessibles par l'utilisateur du système, soit un évaluateur ou un vérificateur?

❑ La piste de vérification générée par le système automatisé peut-elle servir à générer une notification de la décision (y compris un énoncé des raisons ou une autre notification) au besoin?

❑ La piste de vérification est-elle protégée contre la falsification (pour assurer la protection et l'intégrité des données)?

❑ La piste de vérification indique-t-elle précisément quelle version d'un système automatisé a été utilisée pour chaque décision qu'il appuie?

❑ La piste de vérification inclut-elle un historique exhaustif et imprimable des modifications, notamment : qui a créé le dossier (avec l'heure et la date)? Qui a modifié le dossier (avec l'heure et la date)? Qu'est-ce qui a été

20

modifié? Pour les questions de protection de la vie privée et les questions commerciales confidentielles, qui a consulté le dossier (avec l'heure et la date)? Qui a pris la décision définitive (avec l'heure et la date)?

❑ La piste de vérification commence-t-elle par indiquer le pouvoir ou le pouvoir délégué prévu dans la loi?

❑ La piste de vérification indique-t-elle qui est le décideur autorisé?

❑ La piste de vérification permet-elle l'enregistrement de l'intervention humaine dans les processus automatisés, par exemple le nom de la personne autorisée à intervenir?

Regroupez-vous les documents énonçant le développement et le déploiement du système dans son ensemble, y compris les spécifications de la conception, les données de formation, les versions du système et toute autre information qui pourrait être utile pour respecter les exigences de la directive du SCT ou dans le cas d'une vérification, d'une enquête, d'une action coercitive ou d'une procédure judiciaire?

Comment votre approche d'assurance de la qualité continue veille-t-elle à ce que votre système automatisé produise constamment les bons résultats?

❑ Commencerez-vous par veiller à ce que des humains examinent chaque décision passant par le modèle, puis réduirez-vous peu à peu la portion des décisions examinées par des humains?

❑ Votre approche d'assurance de la qualité permettra-t-elle de vérifier s'il y a entrave involontaire de la capacité des agents, préjugés ou discrimination?

Au cours d'un exercice d'assurance de la qualité, vous devriez vous assurer de consigner par écrit votre processus et ses résultats, en donnant assez de détails pour permettre à d'autres de le faire ou de l'examiner en cas de vérification ou de contestation juridique à l'avenir.

Votre système peut-il entraîner des risques liés à la cybersécurité?

❑ Est-il compatible avec les politiques de cybersécurité?

❑ Pouvez-vous nous convaincre que le système est protégé contre le piratage et la manipulation?

Avez-vous désigné un groupe d'experts externes qui peut examiner la conception de votre système de soutien aux décisions automatisé?

Selon l'incidence potentielle de votre projet, la Directive sur la prise de décisions automatisée du SCT pourrait vous obliger à le faire. Même si elle n'est pas obligatoire, la consultation d'experts en IA ou en science des données (par exemple, des experts du Conseil national de recherches) peut vous permettre d'obtenir des suggestions en vue d'améliorer votre système.

Avez-vous élaboré un diagramme de processus détaillé qui permettra aux spécialistes des politiques, aux experts en protection de la vie privée, aux avocats et à d'autres personnes en dehors de votre domaine de comprendre le processus en question et comment, précisément, l'établissement d'un soutien aux décisions automatisé le modifiera?

Pour le SCT :

Avez-vous réalisé une évaluation préliminaire de votre système au moyen de l'outil d'évaluation des répercussions algorithmiques, soulevant particulièrement les exigences liées à ce qui suit?

❑ Intervention humaine

21

- ❑ Mise à l'essai
- ❑ Surveillance
- ❑ Approbation de la mise en marche du système

C. Se préparer au lancement

La présente section répondra à des questions comme celles-ci : *IRCC est-il prêt à mettre en œuvre le nouveau système? Avez-vous informé toutes les personnes concernées et le Ministère est-il prêt à répondre aux questions et à réagir aux interruptions imprévues?*

Avez-vous terminé d'élaborer votre approche concernant la transparence envers le public?

IRCC devrait communiquer aux clients et aux Canadiens tous les cas (secteurs d'activité, fonctions particulières) où le système de soutien aux décisions automatisé est utilisé.

À propos de ces systèmes, vous devriez transmettre au public :

- ❑ Une description du système en langage simple, y compris des exemples de ce qui suit :
    - la raison d'être du système
    - la justification de son utilisation
    - son rôle dans le processus de prise de décisions
    - les résultats possibles pour les gens
- ❑ Une description des données sur lesquelles reposent le système et leur provenance (y compris un lien vers des données de formation anonymes accessibles au public)
- ❑ Une description générale des variables utilisées dans le modèle
- ❑ De l'information sur le rendement, comme les taux d'erreur ou d'exactitude, la répartition des clients touchés (p. ex. par pays ou par programme), les coûts ou la productivité par rapport à une option avec intervention humaine seulement

La communication intégrale du code source au public démontre une transparence élevée, mais la plupart des gens ne sauront pas ce qu'il signifie. Ainsi, ce qui peut être pertinent de révéler au public est différent de ce que l'on divulgue à une communauté d'experts techniques. IRCC devrait divulguer son code source, dans la mesure où la loi le permet, à un groupe d'experts techniques, tout en fournissant au public une explication facile à comprendre sur le fonctionnement du système. Cette approche a l'avantage de protéger le Ministère contre les cyberattaques ou les fraudes sophistiquées perpétrées par des entités malveillantes ayant des compétences techniques, équilibrant ainsi la transparence avec l'intégrité du programme. Le groupe d'experts qui examine le code source et la conception du système de façon générale produisent un rapport contenant une section destinée au public.

- ❑ S'il existe des détails sur le système qui, selon vous, ne devraient pas être divulgués au public, avez-vous confirmé qu'il était possible de protéger cette information? Par exemple, y a-t-il des exemptions convenables aux termes de l'AIPRP ou d'autres mesures de protection législatives, comme la protection pour la sécurité nationale (et, s'il y a lieu, aux termes de la Directive du SCT)?

Avez-vous préparé un avis pour les clients les informant qu'une décision à l'égard de leur demande sera prise en totalité ou en partie par un système automatisé?

Il faut rédiger en langage simple des avis au point de collecte de renseignements personnels et les rendre facilement accessibles sur la page web du programme. La section 6.2.9 de la <u>Directive sur les pratiques relatives à la protection de la vie privée du SCT</u> présente d'autres précisions sur les exigences relatives aux avis sur la vie privée.

IRCC a rédigé l'avis suivant à l'intention des clients dont la demande de visa temporaire serait traitée par l'intermédiaire des projets pilotes d'analytique avancée en Chine ou en Inde :

PROTECTION DES RENSEIGNEMENTS PERSONNELS ET ANALYTIQUE

La protection des renseignements personnels est une considération importante pour le gouvernement du Canada dans le contexte de l'analytique avancée.

Immigration, Réfugiés et Citoyenneté Canada (IRCC) peut utiliser les renseignements personnels fournis dans les demandes et d'autres renseignements fournis à l'appui des demandes pour concevoir un traitement axé sur l'analytique informatisée et peut utiliser l'analytique pour traiter votre demande. Le Ministère procédera conformément à la loi et d'une manière qui cadre avec les fins auxquelles les renseignements sont recueillis.

Les renseignements personnels fournis dans les demandes qui seront utilisés à des fins d'analytique pourraient aussi être utilisés à des fins de recherche, de statistique, d'évaluation de programmes et de politiques, de vérification interne, de conformité, de gestion des risques, d'élaboration de stratégies et de production de rapports.

Avez-vous besoin de modifier les fichiers de renseignements personnels (FRP) pour décrire l'utilisation ou la communication de l'information?

Une stratégie de communication est-elle nécessaire?
- ❑ Le cas échéant, sera-t-elle annoncée?
- ❑ Avez-vous préparé des messages clés, des foires aux questions, etc.?

Dans le cas d'un nouveau système qui changera fondamentalement le travail quotidien des décideurs, avez-vous consulté les Ressources humaines et les syndicats concernés pour déterminer les modifications à apporter aux descriptions de travail génériques, aux énoncés des critères de mérite ou aux classifications de poste?

Les employés touchés ont-ils accès à la formation dont ils ont besoin pour comprendre le nouveau processus opérationnel et interpréter correctement les extrants du système automatisé?

Si les agents interprètent mal l'information que leur fournit un système de soutien aux décisions automatisé (par exemple, en interprétant qu'une étiquette signifiant qu'une demande est « simple » comme une recommandation d'approbation), on pourrait percevoir que le Ministère entrave l'exercice de jugement des agents.

Des instructions sur l'exécution de programmes peuvent être requises pour expliquer les rôles et responsabilités des employés et leur utilisation de l'information provenant de ces systèmes.

Avez-vous réfléchi à la façon dont les besoins en ressources évolueront au fil du temps?
Il est courant d'avoir besoin de plus de ressources humaines et financières dès le départ, au moment où un régime d'assurance de la qualité strict est vital pour veiller au fonctionnement prévu du système. Il est aussi important de prévoir les coûts permanents relatifs à la maintenance et au renouvellement du système.

Les clients souhaitant contester une décision concernant leur demande ont-ils suffisamment de recours pour le faire? La mise en place d'un système de soutien aux décisions automatisé nécessite-t-elle des changements ou des parcours additionnels?

24

Lorsqu'un client conteste une décision qui a été prise par un système de soutien aux décisions automatisé, et qu'IRCC est autorisé à l'examiner et accepte de le faire, le processus d'examen devrait faire en sorte que A) la personne examinant la décision se mette dans la peau du décideur initial, et B) la personne examinant la décision n'est pas contrainte par la décision initiale, ni n'est contrainte de tenir compte toute conclusion ou recommandation fournie par le système informatique qui a contribué à la prise de décision. Si le décideur initial et la personne examinant la décision utilisent la même technologie, il y a un risque qu'un défaut ou une erreur de la technologie se répercute d'un processus à l'autre.

Avez-vous dressé un plan d'urgence dans le cas d'une panne de système (à court terme ou à long terme)?

Avez-vous établi une structure de gouvernance qui définit clairement les rôles et responsabilités des principaux partenaires qui exécuteront le projet, notamment au moment de passer de l'environnement d'essai à un programme permanent?

> La *Directive sur la prise de décisions automatisée* du SCT précise que le sous-ministre adjoint responsable de l'utilisation d'un système de prise de décision automatisée est aussi responsable de l'évaluation des répercussions algorithmiques, des mesures visant la transparence, de l'assurance de la qualité, des recours du client et des rapports.

Avez-vous une stratégie pour que les gestionnaires de projet assurent le suivi des progrès, mesurent les résultats et en rendent compte, et appuient les vérifications ou évaluations ultérieures?

❑ Avez-vous un ou des paramètres servant à mesurer précisément l'incidence des *fonctions de soutien aux décisions automatisées* de votre programme?

Avez-vous respecté toutes les exigences de protection de la vie privée et abordé tous les risques relevés lors de votre évaluation des facteurs relatifs à la vie privée (s'il y a lieu) ou par le Commissariat à la protection de la vie privée?

Avez-vous confirmé auprès des partenaires concernés (par exemple l'ASFC) qu'ils sont prêts pour le lancement de votre système?

❑ Ont-ils besoin de soutien de la part d'IRCC, comme des documents d'information ou des produits de communication?

Pour le SCT :
- ❑ *Avez-vous soumis votre évaluation des répercussions algorithmiques terminée et approuvée au SCT?*
- ❑ *Avez-vous suivi les étapes requises selon les résultats de l'évaluation des répercussions algorithmiques?*
- ❑ *Avez-vous réfléchi aux paramètres que vous utiliserez pour respecter les exigences de production de rapports du SCT?*

* LE PROCESSUS POUR OBTENIR L'APPROBATION DÉFINITIVE DU DÉPLOIEMENT PEUT DÉPENDRE DE LA PORTÉE DES CHANGEMENTS QUE VOUS PROPOSEZ.

LE COMITÉ SUR LA POLITIQUE NUMÉRIQUE ET LE CADRE JURIDIQUE PEUT VOUS INDIQUER LE NIVEAU D'APPROBATION REQUIS (P. EX. SMA, SOUS-MINISTRE OU MINISTRE).

D. Une fois le système fonctionnel

La présente section répondra à des questions comme celles-ci : *Le système fonctionne-t-il comme prévu? Faut-il revoir l'un des facteurs concernés?*

Procédez-vous continuellement à l'assurance de la qualité pour vérifier que votre système ne génère pas d'erreurs ou de résultats accidentels?
- ❑ Réservez-vous une portion suffisante de toutes les demandes à l'examen exclusif d'un agent (humain)?
- ❑ Sélectionnez-vous au hasard des dossiers qui passent par le système de soutien aux décisions automatisé et les mélangez-vous à la charge de travail normale des décideurs humains? Prenez-vous des précautions pour que les agents ne sachent pas quelles demandes sont nouvelles et lesquelles sont des vérifications d'assurance de la qualité?

Vos résultats à ce jour permettent-ils de conclure qu'IRCC pourrait réduire sans risque le seuil de fiabilité du système pour réaliser des gains d'efficacité? Ou l'inverse?

Êtes-vous à l'affût de nouvelles menaces de cybersécurité?

Avez-vous de nouvelles données qui pourraient améliorer la prochaine itération de votre système de soutien aux décisions? *Par exemple, des données qui pourraient remplacer les intermédiaires avec des preuves de résultats plus directes?*

Y a-t-il un cycle et un processus établis pour l'examen périodique des règles opérationnelles, de la qualité des données, des seuils de fiabilité et des gains de productivité?

Sans une approche établie à l'examen périodique, il y a un risque que le fonctionnement adéquat du système repose sur le maintien en poste de certaines personnes. IRCC doit constamment cerner les limites des algorithmes (et des données sous-jacentes) et en informer les décideurs, peu importe le taux de roulement du personnel ou le changement de rôle.

Si des modifications au système deviennent nécessaires, mais ne peuvent être apportées immédiatement, quelles stratégies provisoires appliquerez-vous pour veiller à maintenir l'exactitude de la prise de décisions (par exemple, des alertes dans le système et des avis d'instructions provisoires destinées aux utilisateurs du système)?

4. GLOSSAIRE DES TERMES CLÉS

Algorithme : Séquence d'instructions, de règles et de calculs exécutée par un ordinateur selon un ordre particulier pour obtenir un résultat, habituellement une réponse à un problème donné. Il est possible de combiner plusieurs algorithmes pour résoudre des problèmes complexes.

Analytique des données (y compris l'analytique avancée) : Extraction de signification de grands volumes de données (parfois à haute vitesse) au moyen de systèmes informatiques spécialisés. Ces systèmes transforment, organisent et modélisent les données pour tirer des conclusions et cerner des tendances.

Apprentissage machine : Sous-catégorie de l'intelligence artificielle, l'apprentissage machine désigne les algorithmes et les modèles statistiques qui apprennent à partir d'exemples, de données et d'expériences au lieu de suivre des règles préprogrammées. Les systèmes d'apprentissage machine exécutent efficacement une tâche précise sans suivre d'instructions explicites en s'appuyant plutôt sur des modèles et des suppositions.

Automatisation des processus : Aussi appelée « automatisation opérationnelle » (et parfois même « transformation numérique »), l'automatisation des processus désigne l'utilisation de technologies numériques pour exécuter des processus opérationnels courants dans le cadre du déroulement du travail. L'automatisation des processus peut rationaliser une organisation pour la simplicité et l'amélioration de la productivité en libérant les humains des tâches simples et répétitives et en confiant celles-ci à des machines qui les exécutent plus rapidement. Un large éventail d'activités peut être automatisé, ou plus souvent, partiellement automatisé, où l'on maintient l'intervention humaine à des points stratégiques dans le déroulement du travail. Dans le domaine de la prise de décisions administratives à IRCC, l'« automatisation des processus » diffère du « soutien aux décisions automatisé », la première expression s'appliquant à des tâches administratives simples et la deuxième, à des activités où il faut faire preuve d'un certain degré de jugement.

Boîtes noires : Outils logiciels opaques fonctionnant en dehors de la portée de la surveillance et de la responsabilisation. Il s'agit souvent de systèmes d'apprentissage approfondi. Leur comportement peut être difficile à interpréter et à expliquer, soulevant des préoccupations quant au caractère explicable, à la transparence et au contrôle humain.

Code source : Programme informatique dans son langage de programmation original, lisible par l'humain, avant sa traduction en code objet. Il s'agit d'algorithmes et d'instructions informatiques et peut inclure des commentaires du concepteur.

Données massives : Terme familier désignant les énormes volumes d'information qui inondent les organisations au quotidien. Ces données sont souvent très variées et doivent être traitées au moyen de systèmes informatiques rapides; c'est pourquoi il est impossible de les gérer efficacement et d'en tirer de la valeur opérationnelle sans des outils logiciels et des méthodes d'analyse précis.

« Entrave » de l'exercice du jugement de l'humain : Il y a entrave lorsqu'un décideur n'exerce pas véritablement son jugement indépendant sur une question. Cela peut se produire lorsqu'un décideur se contraint à une règle fixe de politique, l'opinion d'une autre personne ou les résultats d'un système de soutien aux décisions. Bien que la personne qui prend les décisions administratives puisse être favorablement influencée par des considérations politiques et d'autres facteurs, elle doit tenir compte des circonstances propres au dossier en question et ne pas

28

s'accrocher aveuglément à un élément (p. ex. la note relative au risque fournie par un système algorithmique) en excluant d'autres facteurs pertinents.

Équité procédurale : Vitale pour la prise de décisions administratives, l'équité procédurale exige que les demandeurs :

— participent véritablement (ils connaissent le dossier et ont l'occasion de le défendre) et bénéficient d'une évaluation juste et exempte de préjugés de leur demande
— soient tenus informés des preuves négatives et des préoccupations des décideurs
— se voient offrir une possibilité réelle de répondre à de telles preuves et préoccupations à l'égard de leur demande
— se voient présenter les raisons de la décision

Il y a de nombreux facteurs qui déterminent ce qu'il faut aborder précisément pour respecter ces principes. L'un de ces facteurs est l'importance de la décision du point de vue du demandeur.

Gouvernance des données : Efforts déployés à l'échelle de l'organisation pour gérer officiellement les biens de données (par exemple, qui fait quoi, quand et selon quel pouvoir). Grâce à une gouvernance des données efficace, on s'assure que la qualité des données est supérieure au moment où elles sont créées, extraites et manipulées pour la production de rapports ou la prise de décisions. Faisant le pont entre le domaine des opérations et celui des TI, la gouvernance des données appuie l'innovation des processus et l'amélioration du service aux clients.

Intelligence artificielle : Englobant un large éventail de technologies et d'approches, l'intelligence artificielle est essentiellement le domaine des sciences informatiques consacré à la résolution de problèmes cognitifs associés à l'intelligence humaine, comme l'apprentissage, la résolution de problèmes et la reconnaissance de formes.

Prise de décisions administratives : Décisions prises par des responsables autorisés d'une institution (comme un conseil, un tribunal, une commission, un ministère ou un organisme du gouvernement ou un ministre) où les résultats touchent les droits, privilèges et intérêts. Élément central du système d'immigration, la prise de décisions administratives est également essentielle à d'autres domaines, comme le commerce, la fiscalité, la radiodiffusion et le transport.

Règles opérationnelles : Série d'instructions (propositions si-alors) qu'un système automatisé ne peut contourner. Des ensembles de règles peuvent être créés par des humains (selon l'expérience, le raisonnement, la recherche) ou maintenus par des humains (proposés par des modèles d'analytique avancée, mais examinés et approuvés par des humains). Les systèmes automatisés comprennent des règles opérationnelles qui ne peuvent être modifiées sans l'intervention humaine. Ils diffèrent des systèmes *autonomes*, qui sont assortis d'objectifs prédéterminés, mais d'aucune approche prédéterminée pour les atteindre.

Système de soutien aux décisions automatisé : Inclut toute technologie de l'information conçue pour aider directement le décideur humain à prendre une décision administrative (par exemple, en formulant une recommandation) ou pour prendre la décision à la place d'un décideur humain. Il peut s'agir de systèmes comme l'automatisation des prolongations des permis d'études et des fiches de visiteur dans le SMGC ou l'autorisation de voyage électronique.

29

AUTOMATED DECISION SUPPORT
Policy Playbook

How to ensure responsible use of automation, advanced analytics and artificial intelligence in service of administrative decision-making

1. PREAMBLE

Why do we need new policy guidance in this domain?

The premise for automating aspects of administrative decision-making at IRCC is familiar: application volumes are surging, pressure is mounting on our operational network, and technology is seen as our best hope to keep pace and still maintain quality. Traditional business models are simply too resource-intensive for the scale of operations required in 2019 and beyond.

Big data and artificial intelligence (AI) are exploding, and IRCC is beginning to recognize the tremendous potential they bring for new business insights and efficiency gains. But this is a new area, and conventional methodologies, guardrails and governance models are often incompatible with how algorithmic approaches really work. IRCC has thus far acquitted itself reasonably well with an ad hoc approach to resolving novel questions about privacy, transparency, procedural fairness and human-machine interactions. However, with AI and automation (and their associated risks) growing in importance, the Department would benefit from a considered approach. To be successful, IRCC's innovators need clear and readily available guidance, and they need it tailored to our specific operating context.

Treasury Board Secretariat is introducing a new government-wide *Directive on Automated Decision-Making*. The TBS policy, slated to take effect in April 2020, will outline high-level requirements for Departments using systems, tools or statistical models to recommend or make administrative decisions about clients. It will be accompanied by a tool to support Algorithmic Impact Assessments, which will be mandatory before any new system can move from a test environment to live implementation. These overarching policy instruments will help IRCC to navigate uncharted waters, but the guidance is likely too broad to resolve specific issues faced by IRCC system designers.

The TBS Directive also has a different starting point than the internal policy proposed herein. TBS begins with the premise that the reader has already decided to pursue an algorithmic solution and is now in the throes of system development. The policy comes in to make sure that the system, once complete, is up to standard. IRCC's internal policy rewinds the tape and starts with more fundamental questions: *Is automated decision support a good idea in my case? What could I actually automate? Where do I start?*

As a service provider for both Canadians and foreign nationals, IRCC needs guidance that takes into account the Department's unique position. IRCC staff need policy guidance that will ground key technology ethics concepts in our day-to-day work, and directly answer the questions we are asking ourselves. Questions like: *Could we ever use algorithms to automate negative decisions on client applications? When, if ever, are "black box" algorithms appropriate? Who is accountable for individual decisions made by a machine? How do we meaningfully explain automated decisions to clients? How can we give IRCC's officers the benefit of insights gleaned from analytics without fettering their decision-making?*

IRCC needs to find the right answers to these questions. Public confidence is critical to successful immigration, refugee, citizenship, settlement and passport programs, and the public is currently wary of the potential impacts of widespread AI and automation. Codifying or "hard-wiring" business rules, whether based on advanced analytics or on common sense heuristics long relied upon by individual officers, makes them subject to more scrutiny. The application of business rules at tremendous scale multiplies risks that could arise if something goes wrong. A new scale of impact means heightened legal risk for the Department, one of the most heavily litigated in the Government of Canada, and one that already relies on automated systems to deliver its programming. Computerized processing is essential to programs like eTA and Express Entry, and automated decision support could soon become an indispensable component of TR and other program lines.

On top of controlling for anticipated risks, specific policy guidance on automated decision support can help IRCC realize the full benefits of new approaches. Principles speak not only to what actions the Department should avoid because they are too risky, but also to what activities we should pursue because they can *make us better*.

1

What is the scope of activities this playbook targets?

This policy targets **automated systems that directly support or make, in whole or in part, administrative decisions.** This includes systems that: classify cases according to the level of scrutiny they require; flag cases for human review or investigation; provide recommendations about whether applications should be approved; or render complete decisions. The business rules applied by an automated decision support system could be derived from sophisticated data analytics, or from interviews with experienced officers – what matters is that these systems take up a new role in IRCC's decision-making model. Analytics-based rules are a particular concern of this policy, especially if machine learning is involved.

This policy does not cover all automation, all AI or all administrative decision-making. Routine process automation, such as the programming of computers to accomplish repetitive clerical tasks formerly done by humans, is out of scope. So too are systems that merely help to divide our collective caseload between networks, offices or officers for the purpose of efficiency. If the automation does not play a role in helping IRCC to reach an administrative decision that affects a client's fortunes, it is not of primary concern in this policy. Similarly, applications of AI, no matter how advanced, that serve to enhance client service (such as a chatbot) or help the Department manage internal resources, are not the focus here. The focus is on decision-making – more precisely, decision-making that breaks from the traditional model of a single officer reviewing a single set of facts about a single application.

What shape does the playbook take?

Full spectrum support for ethical conduct means having both a code of personal ethics and a set of institutional ethics. On the personal side, nothing new need be developed. The existing Values and Ethics Code for the Public Sector is domain-agnostic (i.e. it already applies equally to accountants doing internal audits as it does to data scientists developing machine learning systems). The code's five core values of respect for democracy, respect for people, integrity, stewardship and excellence provide a sound and enduring basis for IRCC's innovators to make decisions about how to comport themselves ethically and professionally. With respect to institutional ethics, a set of guiding principles is proposed in Part 2.

The principles presented in this policy are designed as a companion to the more concrete how-to guide contained in Part 3. As a general rule, principles do not identify specific courses of action, but instead serve as a basis for decision-making when people find themselves in new territory, confronting novel problems. By disseminating the guiding principles in Part 2, IRCC hopes to see its data scientists, program designers and policy developers inspired, trained and empowered to prioritize ethical considerations in the development of automated decision support systems. In particular, they provide guidance to the Department as a whole and its senior managers, who are ultimately accountable for the results of automated systems and for creating a workplace that fosters responsible use of technology. Reflecting upon and openly discussing the wider impacts of the Department's innovative work can only help to ensure that it consistently supports the public good.

Part 3 offers a handbook, or how-to guide, for policy and program specialists who are considering the development of a new automated decision support system. This section systematically guides innovators through a linear process, encouraging them to consider the right questions at the right times: (1) when determining whether a data-driven solution is well suited to the problem at hand, (2) when setting out to build a new system, (3) when preparing to launch, and (4) once an automated system is up and running.

The document concludes with a glossary of key terms.

> *A new policy on automated decision support is an opportunity to ensure that the Department's thinking keeps pace with the speed of technological change, and that our people and practices continue to deliver a suite of programs equal to the expectations of Canadians and the world. The foundational principles and standards that have guided IRCC are not changing. Our commitments, mission and duties remain consistent. But while it may seem that the tools we use in the digital era merely recreate the paper and in-person practices of the past, they do not. Adoption of these technologies has and will change what we do and how we do it, the skills that we need, our relationship with partners and clients, and the culture of our workplace.*

Note: This policy should be seen as a living document that needs to be regularly updated over time to ensure continued relevance as the technology, and our knowledge thereof, evolves.

2. A SET OF GUIDING PRINCIPLES

How should the Department as a whole proceed when faced with difficult new choices about specific applications of emerging technology? A set of principles is proposed to help IRCC best select for effectiveness, accountability and responsible design.

OVERARCHING GOALS

1. ***The use of new tools should deliver a clear public benefit. IRCC should use automated decision support wherever it can do so responsibly, effectively and efficiently – in that order.***

❖ **IRCC's use of automated decision support should build trust in Canada's immigration system**. Maintaining the confidence of Canadians is critical to the Department's success, and there is significant public anxiety over fairness and privacy risks associated with Big Data and AI. The Department should take measured steps, particularly at the outset, to build legitimacy around its use of automation, analytics and AI. New tools should be designed and implemented in a manner that commands trust and understanding.

❖ There is an opportunity cost to <u>not</u> pursuing the use of automation, advanced analytics and AI. IRCC should not be so cautious as to miss out on the tremendous potential of data-driven technologies. As in any era, **making use of the best tools available is part of effective and efficient use of public money.** In some cases, algorithmic systems are the only efficient way to process the massive amount of information needed for the operation of modern government services.

❖ However, **IRCC can only employ these technologies to the extent that their use remains consistent with Canadian norms and standards** regarding equality and non-discrimination, procedural fairness, transparency, privacy and accountability. The requirements of administrative law, and of overarching frameworks like the Charter of Rights and Freedoms and the *Privacy Act*, are not changed by technology.

❖ **IRCC should not pursue automating decisions (or aspects of decisions) at the expense of program integrity.** Risk management is at the heart of IRCC's work. Automation may change risks or introduce new ones, but innovation should boost program integrity. At worst, it should have a neutral impact in this regard.

❖ Big data and automation should **strengthen administrative decision-making by providing a richness of relevant information to decision-makers, and by focusing their attention on the determinative factors**. In this way, new tools should help officers make not only faster, but *better* decisions.

❖ **IRCC should make every effort to obtain reliable outcomes data** (e.g. from an entry/exit system, from tax records, from provinces and territories) in order to make a direct link between clients' ultimate outcomes and the information they submit at the application stage. At the same time, **the Department needs to carefully consider the strength of the inferences it draws from new data sets**. Differentiating correlation and causation is a perennial challenge, and human intentions are hard to predict reliably.

❖ **Big data should be used to help IRCC look backward as well as forward.** Poorly designed algorithms can introduce or perpetuate bias at scale, but **well-designed models can shine a light on historical bias** by finding

3

previously unidentified patterns in data. IRCC should look to use new tools to detect, assess and remove any bias or inconsistencies that may have inadvertently crept into long-standing business practices.

❖ **IRCC should strive to constantly improve the quality of its data holdings, given that good data is foundational to advanced analytics and AI.** Prioritizing the development of data-driven tools means prioritizing the breadth, depth and reliability of our data sets.

2. *Administrative decisions are about people, and they are made by people, even when we use AI. Humans, not computer systems, are accountable for decisions.*

❖ **The introduction of automated decision support can change the** *time and place* **of human intervention in the decision-making process** (for example, humans can take on new roles, like setting business rules for an automated triage system to later carry out) **but does not displace the central role of human judgment in decision-making.** There is an essential role for humans in deciding which types of systems to use, which cases to apply them to, and which values to encode.

❖ **Keeping a human in the loop is important for public confidence. New business models that incorporate automated decision support should, one way or another, maintain a role for people in the decision-making process.** While human decision-making may not be superior to algorithmic systems built on sophisticated data analytics in all cases, human-in-the-loop systems currently represent a form of transparency and personal accountability that is more familiar to the public than automated processes. In cases where algorithms play a role in administrative decisions that significantly affect people's lives (such as whether someone is granted permanent residence), it is reasonable for applicants and Canadians to expect that a real person has exercised human judgment in developing the process, and that a person is ultimately accountable for its final results.

❖ Automation, advanced analytics and AI are force multipliers; they **assist and augment the capabilities of people**. With ever-increasing intake volumes across IRCC business lines, technology should be oriented toward helping our workforce streamline their efforts. Allowing IRCC staff to focus on more creative, problem-solving, or client-facing work should improve both service delivery and job satisfaction.

❖ **IRCC should endeavour to provide its decision-makers with the best information and tools we have at our disposal.** Putting new insights at the fingertips of decision-makers and focusing their attention on the most relevant factors will eliminate wasted effort, boost consistency and better link action and rationale.

❖ Accountabilities are not passed to machines when they take over tasks previously done by humans. **IRCC must take ownership for the successes and failures of its systems.** Referring to systems or tools as "biased" places the responsibility for negative outcomes on algorithms. This is an outsourcing of moral obligation.

THE RIGHT TOOLS IN THE RIGHT CIRCUMSTANCES

3. *Because IRCC's decisions have significant impacts on the lives of clients and Canadians, the Department should prioritize approaches that carry the least risk.*

4

❖ Automating final decisions is the highest risk approach to achieving efficiencies. Because of this, and because final decisions are rarely the most resource-intensive part of the decision-making process, IRCC should first **explore whether the same efficiencies can be achieved by automating other aspects of the business flow** (e.g. workload distribution, risk-tiering, note-taking, research, communications, etc.) Process automation, such as automated application completeness checks, can "set the table" for decision-makers, so that they can exercise their discretion without first having to clear barriers that could be removed by machines.

❖ **The degree of human involvement in decision-making should be calibrated to:**
   1. the gravity of the decision, including its impact on clients;
   2. the degree of complexity and amount of subjective judgment involved; and
   3. the quality of the data that can be brought to bear.

These considerations are more important than whether the action in question is a final decision or an interim step, or whether it is a positive (approval) or negative (refusal) decision.

❖ When using a predictive model for decision-making, automating approvals is usually considered less impactful than automating refusals (as doing the latter can be viewed as unjust profiling), and this is generally true from the perspective of an individual applicant. However, it is important to view the situation from the perspective of other applicants – particularly one who is applying to a business line where space is limited. **In situations where demand exceeds supply, the stakes for positive decisions are raised**: it may be perceived that an approval (or invitation or acceptance) of one candidate's application comes at the expense of another's.

❖ **A third perspective is that of Canadians en masse**, who accrue the economic, social and cultural benefits of temporary and permanent migration, but who also take on the safety and security risks. **From a collective perspective, automated approvals based on prediction can carry significant consequences**. A system that approves applications without sufficient vetting would raise risks to Canadians, and it is understandable for Canadians to be more concerned about mistakenly approving risky individuals than about mistakenly refusing *bona fide* candidates. IRCC needs to take a balanced view, considering fairness from both the individual and collective perspectives.

4. *"Black box" algorithms can be useful, but cannot be the sole determinant of final decisions on client applications.*

❖ **True "black box" AI tools are of limited utility for administrative decision-making** because the Department should not make decisions it cannot meaningfully explain. "Black box" tools, such as a facial recognition application, can be used in a supporting role – but even then, **confidence/reliability levels should be sufficiently high, and where they are not, humans must have an opportunity to review the system's results before final decisions are made.**

❖ Not all decisions taken by IRCC concern the substantive rights of clients and need to be explained to them. **Black box algorithms should be permitted for the purposes of business optimization**, such as figuring out where to locate offices or how to best distribute resources among them.

RESPONSIBLE DESIGN

5. ***IRCC must recognize the limitations of data-driven technologies and take all reasonable steps to minimize unintended bias.***

❖ Data will always bear the marks of its history. In using data to train a system to make recommendations or decisions, we must be fully aware of the workings of this history. Every machine-learning algorithm operates entirely within the world defined by the data used to calibrate it, so **limitations or flaws in data sets will bias outcomes**, sometimes severely. In creating tools designed to make decisions on immigration, citizenship and passport applications, **IRCC must be careful not to replicate and "hard wire" historical biases** under a guise of technological neutrality. Understanding our data and planning for its responsible use is not a purely technical question – interdisciplinary collaboration is required to address fairness meaningfully.

❖ All analytical tools are limited in their ability to predict and describe outcomes. While overall patterns and trends are useful, they cannot substitute for evidence of personal actions. For this reason, **IRCC should be very careful when applying information about a group to an individual**.

❖ Analytics may suggest that an applicant is low-risk, based on the alignment of their personal profile with that of successful applicants in the past, and in this case **IRCC may wish to risk-manage an automated approval (i.e. the Department takes on the risk). However, the calculus changes in the reverse scenario**: if analytics find that an applicant's profile is consistent with unsuccessful past applicants, suggesting that a refusal may be in order, IRCC should not allow an automated refusal (i.e. the client takes on the risk). As in the case of a humans-only decision-making model, **refusals must be based on *evidence* about the particular person in question, and not solely on their similarity to a general profile of applicants** who have been refused in the past or otherwise indicate riskiness. For example, an automated refusal based on a TRV applicant's passport being in Interpol's Stolen and Lost Travel Documents database would be justifiable because it is based on an objective, verifiable fact about the applicant in question (i.e. they do not meet the requirement for a valid travel document). Conversely, it would not be justifiable to base an automated refusal on the fact that the applicant's country of origin has a high incidence of passport fraud – this could serve as a risk indicator amongst others, but it is not itself a sufficient basis for an outright refusal.

6. ***Officers should be informed, not led to conclusions.***

❖ Decision support systems can be used to assist officers in exercising their discretion. In these cases, systems must be designed so that they **do not fetter decision-makers** in the exercise of their power by pointing them to particular outcomes.

❖ **Officers need appropriate training and context to understand what, precisely, a decision support system is telling them**. For example, labels matter: when a system automatically triages clients' applications into groups, it makes a difference whether these groups are labelled "green/yellow/red" or "straightforward/average/complex." The former may appear to represent instructions to officers (green=approve, red=refuse), the latter a simple indication of how much time and effort an officer will need invest on the way to a decision.

❖ A system that provides **no real opportunity for officers to reflect** is a *de facto* automated decision-making system, even when officers click the last button.

6

❖ It is important to establish a **high level of consistency between how officers input data and how the results of data analytics are presented back to them**. Developers need officers' case notes to be consistent and clear, and officers need to understand what a system means when it provides information to support decision-making.

7. *Humans and algorithmic systems play complementary roles. IRCC should continually strive to optimize these roles and find the right balance, in order to get the best out of each.*

❖ **Let computers focus on their strength** – reliably analyzing large volumes of data and considering millions of possibilities without fatigue – **and let humans focus on theirs**: intuition, creativity, empathy, social skills, shaping a larger strategy. Humans are best at framing questions, machines are best at finding the answers. **Algorithmic systems should free people to focus on things that require their expertise and judgment.**

❖ **Systems that make administrative decisions on their own** – as opposed to assisting a human decision-maker in making a decision – **are suitable for decisions involving straightforward, factual determinations** (such as whether a client has a criminal record or has visited Canada before). Automation should be focused first and foremost on routine elements – where desired outcomes are easily described and where human judgment is not required.

❖ When determining whether a task or step in a decision-making process could be shifted from humans to machines, program designers should ask themselves, "**Is this a situation in which reasonable minds may differ?**" If yes, automation is ill-advised. Conversely, if it is felt that virtually any officer would reach the same conclusion given the facts at hand, automation should be pursued.

❖ **Having a human in the loop not only safeguards against procedural fairness risks, it gets better results.** For straight-forward types of assessments (e.g. photo matching), the performance of data-driven algorithms is often superior to that of humans. But when machines and humans team up, the results are better still. One way to boost performance is to have experts apply judgment to the output of mathematical models; another is to flip the sequence and have program experts and decision-makers contribute to the design of the model in the first place.

❖ **IRCC should regularly review and adjust its automated systems** to ensure they reflect the reality on the ground and are operating in alignment with the policy priorities of the day. Feedback loops are critical to responsible design and quality control. For example, feedback from frontline officers must be able to reach system developers, so that IRCC can constantly reassess whether a system's conclusions can be trusted. Similarly, because automated decision systems cannot rebalance policy priorities on their own, system designers will need to make regular adjustments to capture changes in direction.

❖ Even if IRCC becomes heavily reliant on automated decision support to meet demand, **it is essential that human officers continue to review a portion of applications manually**. This will help to ensure that officer skills do not atrophy and that IRCC's knowledge about local country conditions and fraud trends remains up-to-date.

❖ **Technical experts (and their managers) are in a position of trust, and have a special responsibility to thoroughly evaluate computer systems and their possible impacts.** As always, the Department expects objective assessments and frank advice. With the advent of machine learning systems, extraordinary care should be taken to identify, explain and mitigate potential risks, especially when those risks may only be apparent to those with a technical understanding of the systems and models being used. Managers need to create conditions that allow experts to consult peers, reassess risks as systems evolve and report hazards without fear of reprisal.

7

❖ IRCC should make every effort to **give its employees the skills, tools and support they need** to pursue innovation. Support includes time to learn and space to experiment responsibly.

8. ***On top of respecting Canada's current privacy protection framework, IRCC should continually adopt emerging privacy-related best practices in a rapidly evolving field.***

❖ Because there can be a tension between data-driven technologies and privacy safeguards, data scientists, program designers and IT specialists should be conversant in privacy issues, including **understanding the <u>rights and responsibilities associated with the collection, use, disclosure and retention of personal information</u>**, as well as **emerging best practices and standards regarding the use of other information related to individuals**. IRCC should give thought to mandatory training, as it may be insufficient to let staff self-identify their training needs relating to privacy in the age of big data.

❖ IRCC shares facts about individual clients with certain domestic and international partners, and the Department may wish to share information about its automated systems, including in some cases the algorithms themselves. However, as a general rule, **IRCC should not share system-generated predictions (such as risk scores/flags/recommendations) about individual clients unless: (A) sharing is required by law (e.g. in accordance with the Security of Canada Information Sharing Act), (B) we share the full context, including a basis for partners to understand and assess the merits of the statistical conclusion, and (C) there is a way for the information to be pulled back, corrected or modified as circumstances change.** This is a safeguard in the event of inadvertent algorithmic bias. If such bias unfairly disadvantages an individual, extensive sharing amongst partners can produce a cascading effect, multiplying the discrimination experienced by that person. Exceptions to the no-sharing rule should be carefully considered and approval should be sought at the Deputy Minister level.

❖ The Department should **be proactive in preparing for the possibility of a data breach** related to one or more of its algorithmic systems. This means having the right team and standard operating procedures in place.

TRANSPARENCY AND EXPLAINABILITY

9. ***IRCC should subject all systems to ongoing oversight, to ensure they are technically sound, consistent with legal and policy authorities, fair and functioning as intended.***

❖ Inviting **external scrutiny** of the design of our systems is an important aspect of ensuring legitimacy. Outside experts can help to ensure that systems are both technically sound and free of unintended bias. Where possible, new algorithmic systems should be made available to external experts, such as an advisory body, for auditing, testing and review. If testing is kept internal, the methods and assumptions used in testing, along with the results, should be openly documented and made available to the advisory body.

10. ***IRCC must always be able to provide a meaningful explanation of decisions made on client applications.***

8

❖ Regardless of the degree to which IRCC relies on algorithmic systems to support decision-making, applicants will have the same rights they have today to understand the basis of decisions made on their applications. Explanations are key to applicants' ability to pursue recourse mechanisms, and to government accountability more generally (Parliamentarians and journalists also reasonably expect a thorough explanation of how IRCC makes decisions). Explaining how algorithmic systems work can be technically challenging, but **explanations of automated decisions should not hinge on the general public's understanding of how algorithmic systems function**. If an explanation provides little meaningful information, it offers little value.

❖ One way to make explanations understandable to humans is to limit a system's performance to those functions that can be easily explained. This may, however, mean squandering the potential of AI systems that can make sense of complex data in ways that humans cannot. **Rather than imposing sharp limits on what new tools can *do*, IRCC should endeavour to find new and better ways to *explain* them.**

❖ Given a prohibition on black box systems for administrative decision-making in full, the use of automated decision support systems should make decisions **more explainable, not less**. For example, audit trails capturing every factor and step in the decision-making process can be used, where needed, to systematically generate and record reasons for acceptance or refusal.

❖ An **audit trail** is critical to respecting an applicant's right to understand the basis of the decision on their application – whether positive or negative – and may be needed to satisfy the courts in the context of a legal challenge. Any new tool that lacks the capacity to **automatically generate a comprehensive audit trail of the administrative decision-making path** is not yet ready for implementation.

❖ As with explanations of individual administrative decisions, **IRCC needs to find ways to be meaningfully transparent about the workings of whole systems.** While strategies for explaining the Department's work may differ depending on the audience, at a minimum:
  — Clients need to understand how their applications will be processed and their personal information used
  — Canadians need to understand how IRCC spends public money to facilitate the entry of family, friends and legitimate visitors, while at the same time protecting against threats to health, safety and security
  — Stakeholders, media and technical experts need a basis for an informed critique of Canada's approach

11. IRCC must balance transparency with the need to protect the safety and security of Canadians.

❖ The Department has an interest in not disclosing too much information about its decision systems for program integrity reasons. Divulging too much information can allow individuals to manipulate decision-making, or "game the system." Explanations of how systems work, and of the individual decisions they help the Department to make, should be **detailed enough to inform a client generally about how decision will be made, but not so comprehensive as to jeopardize program integrity and, by extension, the safety and security of Canadians.**

❖ Given that all automated decision-making rules could ultimately find their way into the public domain, **IRCC should endeavour to design systems, to the greatest extent possible, around factors/data points that are difficult to falsify or misrepresent.** For example, bank records are much more difficult to fabricate than a reason for travel, which is simply stated by the applicant. As much as we can, we should rely on objective facts, in order to mitigate the risk of "gaming by claiming." The weight assigned to different data elements in IRCC's algorithms should account for the degree to which they may have been targeted for fraud in the past, or are likely to be targeted going forward.

9

12. Clients will continue to have access to the same recourse mechanisms, and IRCC's use of automated systems should not diminish a person's ability to pursue these avenues.

❖ When a client challenges a decision that was made in full or in part by an automated system, and IRCC has the authority and agrees to review the decision, **a human officer will conduct the review**.

❖ However, **IRCC should not proactively offer clients the choice to have a human officer review and decide on their case in lieu of the automated system at the beginning of the application process**. IRCC's objective is to implement responsible, effective and efficient systems. Allowing clients to opt out of them would compromise the Department's ability to ensure consistent and fast processing, while also erroneously implying that its automated systems provide service inferior to that provided solely by human officers.

3. THE AUTOMATOR'S HANDBOOK

A. Exploring automated decision support as a solution

This section will help to answer questions like: *Does the problem you are trying to solve lend itself to a data-driven solution? Do you have what you need to be successful? Do the benefits outweigh the risks?*

Are you planning simple process automation, or will your proposed intervention contribute, in any way, to decisions about whether to approve client applications?

Some steps in a workflow, such as triage, may seem removed from final decisions, but in reality they may be partly determinative of an approval or refusal. For example, if an automated system sorts incoming applications into three bins – green for straightforward, yellow for average and red for complex – officers may view the green/yellow/red labels, consciously or unconsciously, as recommendations about the riskiness of applications and, therefore, about which they should approve.

Policy and program specialists aiming to automate any part of a decision workflow should pause to reflect on whether their proposed automation is purely process oriented – i.e. it would computerize mundane repetitive tasks formerly done by humans – or whether it could amount to something more.

IF PURE PROCESS AUTOMATION ⟶ SUBSEQUENT SECTIONS OF THIS POLICY DO NOT APPLY
Add your project to **this list** of automated systems used by IRCC

IF POTENTIALLY A CONTRIBUTOR ⟶ KEEP READING
TO DECISION-MAKING

Are you proposing a system that would make direct recommendations to officers about whether to approve or refuse an application, or one that would reach final decisions without human intervention?

NO ⟶ KEEP READING
THIS POLICY STILL APPLIES TO AUTOMATED SYSTEMS THAT
SUPPORT DECISION-MAKING IN OTHER WAYS

YES ⟶ KEEP READING
YOU WILL ALSO BE SUBJECT TO THE REQUIREMENTS OF THE TBS
DIRECTIVE ON AUTOMATED DECISION-MAKING

Are you planning a system that uses machine learning to autonomously generate and adjust business rules without human intervention?

NO ⟶ KEEP READING
THIS POLICY STILL APPLIES TO STATIC RULES-BASED SYSTEMS

YES ⟶ KEEP READING, AND READ CAREFULLY
YOUR PROJECT WILL LIKELY REQUIRE SPECIAL ATTENTION AND
TRIGGER THE STRICTEST REQUIREMENTS OF THE TBS *DIRECTIVE
ON AUTOMATED DECISION-MAKING*

11

If automated decision-making is the goal, have you thought about whether simple process automation or new approaches to workload distribution could achieve similar efficiencies? Have you exhausted straightforward options before moving into the more complex and contentious world of automated decision support?

Will you require new policy or legal authority?

The Department has broad legal authority to use electronic tools to deliver on its mandate (see section 4.1 of IRPA and amendments to section 2.2 of the Canadian Passport Order), but you could still require new authorities if your planned automation would entail substantive changes to program requirements (e.g. eligibility criteria), to program outcomes (e.g. which clients are approved), to the collection, use or disclosure of personal information, or to the role of officers in the decision-making process.

You may need to work with policy leads to potentially seek new authorities through a Memorandum to Cabinet or other means.

Have you thought about who needs to be engaged in the discussion?

You will need to assemble a cross-sectoral, multi-disciplinary group, likely including:

IPG and the processing networks	First-hand knowledge of the decision-making process, the role of officers, costs, potential efficiencies, scalability, etc.
Operations Planning and Performance *Data Governance team and Advanced Analytics Lab*	Help assessing data quality/suitability and expertise in applied data science
Strategic Policy and Planning *Digital Policy team*	Advice on policy and legal authorities, responsible design, connections with other initiatives in the Department, the *TBS Directive on Automated Decision-Making*, etc.
A program policy branch, depending on the use case	Possible policy/program changes to complement your initiative
Legal Services	Advice on program requirements, privacy, Charter, administrative law, intellectual property and litigation risks
Digital Strategy Branch (Transformation and Digital Solutions Sector)	How your initiative fits with IRCC's broader technology strategy
Projects Branch (Transformation and Digital Solutions Sector)	Responsible for managing and prioritizing intake of IM/IT projects
ATIP and Accountability *ATIP division*	Advice on whether you need to undertake or revise an existing Privacy Impact Assessment (PIA), privacy notice or Personal Information Bank (PIB)
Communications	Determine whether there is a need to do any public communications regarding your project
Client Experience **For proposals that could directly affect interactions with clients*	Advice on user experience design and testing

Have you thought about whether you need involvement from CBSA and/or another government department?
Although this policy targets IRCC activities, your project could have impacts on partners (who may also have their own policies/approaches regarding analytics, AI and automation).

Does IRCC have the breadth and quality of data required to make the proposal workable?

— Is there enough high-quality, structured data to support reliable inferences? What are the limitations of the data? Is it accurate, up-to-date and truly representative, and does it contain historical biases that could be perpetuated? *You will need to get an assessment of the quality of your data set, with leadership from the departmental data governance function and collaboration between business and IT. This assessment should be approved by a data governance management table, such as the Data Executive Steering Committee.*

— How many of your data points are proxies for the actual criteria or outcomes you are targeting? All of them? *You will need tried and tested indicators that applicants comply with program objectives/requirements.*

— Can you use the data? Does it invade or pose significant risks to individuals' right to privacy? On what basis would you use the data underpinning your analytics in a real world application? Do you only need access to information that IRCC already collects (cross program uses of information previously collected for one program should be reviewed), or are you seeking to pull in external data, such as that held by federal, provincial or international partners? *You will need to get advice from the ATIP division and possibly Legal Services about whether your proposed collection, use or disclosure of the data is consistent with the privacy framework governing the Department, or whether new authorities should be sought.*

Is this a project that requires special attention?

❑ Is it within an area of intense public scrutiny (e.g. because of privacy concerns) and/or frequent litigation?

❑ Are clients in this line of business particularly vulnerable?

❑ Are the stakes of the decisions very high?

❑ Are you thinking of doing automated refusals?

❑ Will it have a major impact on IRCC staff, either in terms of their numbers or their roles?

❑ Are you thinking of using a "black box" algorithm?

❑ Are you thinking of collecting/using non-traditional and potentially sensitive data, such as data associated with clients' social media accounts?

How do you know if you have a "black box"?

❑ Is the predictive algorithm protected from disclosure (e.g. a trade secret)?

❑ Is the process not (precisely) described?

❑ Is the system's process of analysis or reasoning unknowable? Its behaviour difficult to interpret or explain?

If you answered yes to one or more, your project will require more consultation and oversight.

What impacts, positive and negative, will the proposed automation have on clients? *For example:*

❑ Will it lead to faster processing?

❑ Will it require clients to entrust IRCC with more personal information, or otherwise affect their privacy?

What impacts, positive and negative, will the proposed automation have on the Department? *For example:*

❑ Will it save money? (In the short, medium or long term?)

❑ Will it improve consistency?

❑ Will it improve program integrity?

❑ Will it increase the risk of litigation?

❑ Will it change the core competencies, job descriptions or classifications of IRCC decision-makers? Will it change the overall size or make-up of the staff required to deliver a program?

Have you considered how frontline staff might react?

You may wish to consult with them, and with Human Resources. Even proposals that do not entail substantive changes to officer roles can inadvertently provoke uncertainty and trepidation. Efforts to communicate project goals up-front could reassure affected staff, saving time and effort later.

Have you considered how external stakeholders might react?

It will not be practical to consult external stakeholders on the development of every algorithm or automated system used by IRCC, but it is important to consider the perspectives of those who will be affected by automating aspects of the decision-making process, and particularly when new algorithms/systems are part of an entirely new program or a substantial change in delivery approach.

Traditional IRCC stakeholders are still important, but you will likely need to consider a broader array of stakeholders, such as those working in the field of AI/technology ethics.

It is recommended that you document stakeholder perspectives as you would when developing a significant policy or legislative change.

Have you considered whether you will need to undertake a <u>Privacy Impact Assessment</u>, or revise an existing one? The ATIP division can help you make this determination.

Have you taken advantage of available training related to data-driven technologies and their effects?

Have you modeled more than one option?

It is unlikely that there is only one way to achieve your goal. Exploring, mocking up and testing some alternatives might confirm your initial hypothesis and strengthen your business case – or it might help you to uncover an even better approach. Remember that the privacy framework applies even in the modeling stage of your activities.

Has your modeling generated new insights about how your program could be designed?

For instance, the expression of interest system facilitated by Express Entry has highlighted opportunities to modernize the entire federal economic program. As well, the data analytics underpinning the China TRV predictive model has bolstered the rationale for the CAN+ program (i.e. travel history).

— Will your proposed solution effectively replace program requirements established in legislation or regulations with algorithmically-generated business rules?

Have you adequately budgeted for data analytics experimentation and iterative systems development in your Memorandum to Cabinet and/or Treasury Board Submission (if required)?

Uncovering new business insights through the use of advanced analytics takes time and costs money. Building a new decision support system can be a process of trial and error, and budgeting should provide the flexibility to iterate as needed before achieving a final product. Once built, automated systems require resources for maintenance and periodic renewal.

B. Designing your system

This section will help to answer questions like: *How do we design for fairness? How do we safeguard the impartiality of our decision-makers? How do we meet client expectations for transparency and protection of personal information?*

What steps will you take to minimize unintended bias?

Poorly designed algorithms can introduce bias, but in most cases bias comes from the data that is used for analytics or for the training of machine learning models. Program designers need to be aware of weaknesses in the data at the root of their project. Some techniques for mitigating bias in data include consulting subject matter experts to ensure the data is situated in the appropriate context (e.g. has IRCC changed any business definitions over time?), and subjecting machine learning models to adversarial training that could correct for perceived unfairness. Rigorous quality assurance testing of results can also detect trends that suggest bias, allowing for immediate correction.

Whatever the specifics of a given project, it is always advisable to assemble a diversity of views and identities within your development team. There are numerous examples of AI projects gone wrong due to a lack of diversity within the brain trust. Overlooking the unique circumstances faced by different groups of people can have damaging impacts on the populations your system is designed to assist. Guarding against unwanted bias means bringing a GBA+ lens to your data analysis and early design work. *You should endeavour, to the greatest extent possible, to have team members who bring different worldviews, be they based on sex, gender, age, ethnicity, language, rurality, ability or other identity factors.*

Considering the need for transparency, particularly in the context of legal challenges, are you proposing business rules that are defensible and difficult for individuals to "game"?

— Models and rules need to address program requirements such that decisions are based on these requirements and on relevant, reliable information.

— Are you designing your system, to the greatest extent possible, around factors/data points that are difficult to falsify or misrepresent? Much of the information on client applications is simply stated by applicants, with no way for IRCC to authenticate, but other data points, such as bank records or previously issued visas, are verifiable. Preference for the latter will help mitigate the risk of "gaming by claiming" on the part of applicants who start out with knowledge of how IRCC's systems work.

— If weighting variables in an algorithm, does your weighting account for the degree to which each variable may have been targeted for fraud in the past, or is likely to be targeted in the future?

How will your system reflect Privacy by Design principles?

— Are you considering privacy at every stage of program development?

— Have you proactively embedded privacy in the design of your system architecture?

Your approach to using personal information for legitimate ends, without violating the rights of individuals or groups, should include:

— Understanding the provenance of the data

> **Did you know?**
> You can consult the Government Advisory Directorate at the Office of the Privacy Commissioner (OPC) for advice on privacy matters, even if you are not undertaking a Privacy Impact Assessment. To do so, start by contacting ATIP and Accountability Branch at IRCC.ATIPInternal-InterneAIPRP.IRCC@cic.gc.ca. IRCC's Chief Privacy Officer is the central point of contact for dealing with the OPC.

— Ensuring its accuracy

— Reducing unnecessary, redundant or marginal data

— Taking precautions to prevent re-identification of anonymized (or pseudo-anonymized) data or unauthorized data collection

— Protecting data from unauthorized access and accidental disclosure

Transparent data policies and privacy procedures should allow individuals to:

— Understand what data is being collected, how it is being used and disclosed, and how long it will be kept

— Give informed consent for automatic data collection, and for any after-the-fact data sharing

— Review, obtain and correct inaccuracies in their personal data

How will you design the human-machine interaction?

— Will your system replace human decision-makers, or only assist them?

— If it is an assist tool, what steps will you take to safeguard the impartiality of decision-makers? For example, you must ensure they have a clear understanding of what exactly a decision support system is telling them. System outputs will need to come across as expert advice, and not ready-made decisions (making the machine the de facto decision-maker).

— How does your design maximize the value that humans can add? For example, will humans review all complex cases?

— In what circumstances will a human be able to override a decision made by the system?

Will your system use open source software?

The TBS Directive on Management of Information Technology requires departments to use open standards and open source software by default, with exceptions for circumstances where an open source option is not available or does not meet user needs.

If not using open source, departments are expected to favour platform-agnostic commercial off-the-shelf (COTS) options over proprietary COTS, avoiding technology dependency and allowing for substitutability and interoperability.

If developing a custom-built application, by default any source code written by the government must be released in an open format via Government of Canada websites and services designated by TBS.

How will your system be designed to ensure procedural fairness?

Procedural fairness requires that applicants have a *meaningful opportunity to present their case before an impartial decision-maker*. Exactly what procedural fairness requires in practice varies, but it can include things like:

— being provided with the opportunity to make submissions

— receiving disclosure of negative evidence, and being afforded an opportunity to comment on it

— receiving reasons for a negative decision

IRCC uses a sliding scale of actions – the greater the stakes of a decision, the greater the opportunity afforded clients to respond to concerns before a negative decision is rendered.

In order to preserve (or enhance) procedural fairness, you will need to think carefully about how the addition of automated decision support will change application processing and decision-making. For instance:

❑ If you are planning a predictive model, how will it inform officers without fettering their discretion?

❑ Are all relevant factors still being considered?

❑ Will your system draw on information beyond what an applicant submits? Clients have a right to know what other ("extrinsic") evidence was used in decision-making, in part so they have the opportunity to correct any erroneous information.

❑ What reasons will be provided when a decision is made?

Ensuring meaningful explanations

Procedural fairness means that applicants have the right to understand the basis of the decision made on their application. This is key to their ability to pursue recourse mechanisms, and to government accountability more generally.

IRCC has yet to pursue automated refusals of client applications. If the Department were to do so, program designers would need to be mindful of the following:

— Administrative decisions made by, or with the assistance of, an automated decision support system must be at least as explainable as the decisions traditionally made by humans alone. This is why "black box" algorithms, like advanced neural nets, are not recommended as a primary way of determining whether to approve an application.

— Explanations must not jeopardize program integrity. Fortunately, a level of transparency that would include disclosing the inner workings of an AI system is generally unnecessary. Instead, explanations are mainly required to answer questions like: *What was the basis of a decision? What were the main factors? Would changing a certain factor have changed the decision? Why did two similar-looking cases lead to different decisions?*

— As a general rule, explanations of automated decisions should: (1) help clients understand why a particular decision was reached, (2) provide grounds to contest the decision should the client wish to do so. Program designers should also consider whether it is possible to indicate, at a high level, what would need to change in order for the client to achieve a desired result in the future. This is sometimes called a "counterfactual explanation."

*At present, IRCC gives explanations like, "I am not satisfied that you will leave Canada at the end of your stay as a temporary resident...based on your current employment situation" or "...based on your travel history."

Who will be the decision-maker of record?

<u>IRPA Section 4.1</u> specifies that officers can make decisions, determinations or examinations using automated systems made available by the Minister (similar authorities are being sought for the Canadian Passport Order, and they may also be needed in the Citizenship Act). This does not mean that systems themselves become the decision-maker of record. Humans design system architectures, business rules, confidence thresholds and the like, and humans are ultimately accountable for every decision the Department makes. The Minister, in turn, is accountable to Parliament for all of the Department's activities.

For the China TRV advanced analytics project, IRCC elected to assign the Director General of Immigration Program Guidance Branch as the decision-maker of record for all eligibility decisions made without officer review.

As a general rule, when using a predictive model or automated system to support administrative decision-making, it is logical to demonstrate accountability by making the decision-maker of record the person who signs off on the business rules that the system applies.

Rules should be described in a clear narrative fashion so that there is no question whether the designated decision-maker fully understood the rules and, by extension, the basis of decisions they will underpin.

— Who will approve the business rules that are used in your decision support system?

— Is this the same person who will be logged as the decision-maker of record?

— If an automated system determines some, but not all, elements of an administrative decision, will the person making the final determination be considered the decision-maker of record for the entire decision? Or will

certain sub-elements of the overall decision be treated as separate decisions with their own decision-maker of record (e.g. eligibility vs admissibility)?

How will your system keep an audit trail for each decision?

To prepare for potential legal challenges, it is essential that IRCC establish a chain of evidence about how whole systems work, as well as a trail relating to each individual decision that goes through them.

❑ Does the automated system have the capacity to automatically generate a comprehensive audit trail of the administrative decision-making path, or will you retain an electronic record in the applicant's file?

> For eTA and Visitor Record and Study Permit Extensions, the automation tab in GCMS allows users to clearly see all sub-activities, including when applications "passed" or dropped out for manual review.

❑ Are all the key decision points identifiable in audit trail?

❑ Are all the key decision points within the automated system's logic linked to the relevant legislation, policy or procedure?

❑ Are all decisions recorded and accessible by the system's user, a reviewer or an auditor?

❑ Can the audit trail generated by the automated system be used to help generate a notification of the decision (including a statement of reasons or other notification) where required?

❑ Is the audit trail secure from tampering (to provide protection and data integrity)?

❑ Does the audit trail identify precisely which version of an automated system was used for each decision it supports?

❑ Does the audit trail include a comprehensive and printable modification history, including: who created the record (with time and date)?; who has modified the record (with time and date)?; what was modified?; for privacy and commercial-in-confidence matters, who has viewed the record (with time and date)?; who made the final decision (with time and date?);

❑ Does the audit trail start by identifying the authority or delegated authority identified in legislation?

❑ Does the audit trail show who the authorized decision-maker is?

❑ Does the audit trail enable the recording of human intervention in automated processes, for example recording who is authorized to exercise intervention?

Are you assembling documentation outlining the development and deployment of the system as a whole, including design specifications, training data, system versions and any other information that could be valuable in meeting TBS Directive requirements or in the event of an audit, investigation, enforcement action or judicial proceeding?

How will your approach to ongoing quality assurance ensure that your automated system consistently produces the right outputs?

❑ Will you start by having humans review every decision that is put through the model, and then gradually scale back the share of decisions that are human-reviewed?

❑ Will your QA approach test for unintended fettering of officers, bias and discrimination?

In the course of a QA exercise, you should make sure to document your process and its results, providing enough detail to allow others to repeat the exercise or to review it if an audit or legal challenge occurs in the future.

Could your system create any cybersecurity risks?

❑ Is it compliant with existing cybersecurity policies?

❑ Can we be confident that the system is protected against hacking and manipulation?

18

Have you identified a group of external experts who can review the design of your automated decision support system?

Depending on the potential impact of your project, the *TBS Directive on Automated Decision-Making* may require you to do this. Even if not compulsory, consultation with experts in AI or data science (for example, from the National Research Council) may yield helpful suggestions for an improved system.

Have you developed a detailed process map that will enable policy specialists, privacy experts, legal counsel and others outside of your domain to understand the business process in question, and how, specifically, the introduction of automated decision support will change it?

For TBS:

Have you done a preliminary assessment of your system using TBS's Algorithmic Impact Assessment tool, noting in particular the requirements related to:

- ❑ Human-in-the-loop
- ❑ Testing
- ❑ Monitoring
- ❑ Approval for the system to operate

C.   Preparing for launch

This section will help to answer questions like: *Is IRCC truly ready to implement the new system? Have you informed all the right people and is the Department prepared to respond to questions and unforeseen disruptions?*

Have you finalized your approach to public transparency?

IRCC should disclose to clients and Canadians all instances (lines of business, specific functions) in which automated decision support systems are used.

Public disclosure about systems should include:

- ❑ A plain language description of the system, including examples
    - the purpose of the system
    - the rationale for its use
    - its role within the decision-making process
    - possible outcomes for people
- ❑ A description of the data underpinning the system and where it came from (including a link to any anonymized training data that is publicly available)
- ❑ A general description of the variables used in the model
- ❑ Performance information, such as accuracy/error rates, distribution of affected clients (e.g. by country, by program line), cost or productivity vs humans-only alternative

Fully disclosing source code to the public is highly transparent in spirit, but most people cannot make sense of code. Thus, what can be meaningfully revealed to the general public is different from what can be shared with a community of technical experts. IRCC should disclose source code, where legally permissible, to a panel of technical experts, and at the same time provide a plain language explanation of the system's workings to the public at large. This approach has the added benefit of safeguarding the Department from cyber-attacks or sophisticated fraud on the part of malefactors with technical skills, thereby balancing transparency with program integrity. The expert panel reviewing the source code and broader system design should produce a report that includes a section suitable for public consumption.

- ❑ If there are any system details that you feel should be protected from public disclosure, have you confirmed whether this information can be protected? For example, are there suitable exemptions under ATIP or other legal protections available, such as for protection of national security (and, if applicable, under the TBS Directive)?

Have you prepared a notice for clients, alerting them to the fact that a decision on their application will be made in whole or in part by an automated system?

Notices at the point of collection of personal information should be written in plain language and easily findable on the program webpage. Section 6.2.9 of the TBS Directive on Privacy Practices provides additional detail about the requirements for privacy notices.

IRCC developed the following notice for clients whose TRV applications would be processed through the Advanced Analytics pilots in China and India:

PRIVACY AND ANALYTICS

Privacy protection is a primary consideration for the Government of Canada in the use of advanced analytics.

Immigration, Refugees and Citizenship Canada (IRCC) may use the personal information collected on an application and other information collected in support of an application to design computer analytics processing and may use analytics to process your application. The personal information is used for analytics in accordance with the law and in a manner consistent with the purposes for which the information was collected.

Personal information on an application used for analytics may also be used for research, statistics, program and policy evaluation, internal audit, compliance, risk management, strategy development and reporting.

Do you need to amend any Personal Information Banks (PIBs) in order to describe use or disclosure of information?

Is a communications strategy needed?

- ❑ If so, will there be an announcement?
- ❑ Have you prepared key messages, Qs&As, etc.?

In the case of a new system that will fundamentally change the day-to-day work of decision-makers, have you consulted with Human Resources and the relevant labour union(s) to determine modifications to generic work descriptions, statements of merit criteria, or position classifications that may be required?

Do affected employees have access to the training they need to understand the new workflow and interpret the automated system's outputs correctly?

If officers misinterpret the information given to them by automated decision support systems (for example, by equating a label of "straightforward" with a recommendation to approve), the Department could be seen as fettering its officers' discretion.

A Program Delivery Instruction may be needed to explain the roles and responsibilities of employees and their use of information coming from these systems.

Have you considered how resource requirements will change over time?

It is common to need more human and financial resources at the outset, when a rigorous quality assurance regime is vital to ensuring the system is functioning as intended. It is also important to forecast ongoing costs related to system maintenance and renewal.

Are there sufficient recourse avenues for clients who wish to challenge the decision on their application? Does the introduction of automated decision support necessitate any changes or additional pathways?

When a client challenges a decision that was made with automated decision support, and IRCC is authorized and agrees to review it, the review process should be such that (A) the decision reviewer stands in the shoes of the original decision-maker, and (B) the reviewer is not bound by the original decision, nor are they bound to take into account any findings or recommendations offered by the computer system that the original decision-maker relied upon. If the same technology used by the original decision-maker is also used by the decision reviewer, there is a risk that a flaw or error in the technology will carry over from the initial decision to the review.

Have you devised a contingency plan in the event of a system outage (short-term and long-term)?

Have you worked out a governance structure that sets clear roles and responsibilities for the main partners delivering the project, including when it transitions from pilot to steady state?

Do you have a strategy to ensure that project managers can track progress, measure and report on results, and support future audits or evaluations?

> The TBS *Directive on Automated Decision-Making* specifies that the Assistant Deputy Minister responsible for the program using an automated decision system is responsible for algorithmic impact assessments, transparency measures, quality assurance, client recourse, and reporting.

❑ Do you have a metric(s) to measure specifically the impact of the *automated decision support function(s)* of your program?

Have you met all privacy requirements and addressed any risks that were identified in your Privacy Impact Assessment (if applicable) or by the Office of the Privacy Commissioner?

Have you confirmed with affected partners (e.g. CBSA) that they are prepared for the launch of your system?

❑ Do they require any support from IRCC, such as briefing materials or communications products?

For TBS:

❑ *Have you submitted your finalized and approved Algorithmic Impact Assessment to TBS?*
❑ *Have you taken the steps required by the results of the Algorithmic Impact Assessment?*
❑ *Are you thinking about the metrics you will use to fulfill TBS reporting requirements?*

---

*THE PROCESS FOR GETTING FINAL APPROVAL TO LAUNCH MAY DEPEND ON THE SCALE OF THE CHANGES YOU ARE PROPOSING.

THE DIGITAL POLICY AND LEGAL FRAMEWORK COMMITTEE CAN PROVIDE DIRECTION AS TO THE LEVEL OF SIGN-OFF THAT IS REQUIRED (e.g. ADM, DM, MINISTER).

D. Once up and running

This section will help to answer questions like: *Is the system still functioning as originally intended? Do any intervening factors point to the need for a review?*

Are you conducting ongoing quality assurance to verify that your system is not generating errors or unintentional outcomes?

❑ Are you reserving a significant share of overall applications for human-only review?

❑ Are you randomly selecting cases that go through the automated decision support system and mixing them in with human decision-makers' normal caseload? Are you taking care to make sure that officers don't know which applications are new and which are quality assurance checks?

Do your results to date suggest that IRCC could safely lower the system's confidence threshold in order to realize greater efficiencies? Or the reverse?

Are you monitoring for new cybersecurity threats?

Has any new data become available that could improve the next iteration of your decision support system? *For example, data that could replace proxies with more direct evidence of outcomes?*

Is there an established cycle and process for periodic review of the system's business rules, data quality, confidence thresholds and productivity gains?

Without an established approach to periodic review, there is a risk that the proper functioning of a system will depend on certain key individuals staying in their same job. IRCC needs to identify limitations of algorithms (and the data underpinning them) consistently to decision-makers, regardless of staff turnover or changing roles.

If amendments to the system become necessary but cannot be made immediately, what interim strategies will you employ to ensure that decision-making remains accurate (for example, alerts on the system and notification of interim instructions to system users)?

4. GLOSSARY OF KEY TERMS

Administrative decision-making: Decisions made by authorized officials of an institution (such as a board, tribunal, commission, government department, agency or Minister), where the outcome affects legal rights, privileges or interests. Central to the immigration system, administrative decision-making is also critical in other fields like international trade, taxation, broadcasting and transport.

Algorithm: A sequence of instructions, rules, and calculations executed by a computer in a particular order to yield a result, typically an answer to a specified problem. Multiple algorithms can be used in combination to solve complex problems.

Artificial intelligence: Encompassing a broad range of technologies and approaches, AI is essentially the field of computer science dedicated to solving cognitive problems commonly associated with human intelligence, such as learning, problem solving, and pattern recognition.

Automated decision support system: Includes any information technology designed to directly support a human decision-maker on an administrative decision (for example, by providing a recommendation), and/or designed to make an administrative decision in lieu of a human decision-maker. This includes systems like eTA or Visitor Record and Study Permit Extension automation in GCMS.

Big data: A popular colloquial term used to describe immense information assets that inundate organizations on a day-to-day basis. This data is often highly varied and needs to be processed through computer systems at a fast pace, making it impossible to effectively manage and yield business value without specific software tools and analytical methods.

Black Box: Opaque software tools working outside the scope of meaningful scrutiny and accountability. Usually deep learning systems. Their behaviour can be difficult to interpret and explain, raising concerns over explainability, transparency, and human control.

Business rules: A series of instructions (if-then propositions), from which an automated system is unable to deviate. Rule sets can be human-crafted (based on experience, common sense, research) or human-curated (proposed by advanced analytics models but reviewed and approved by people). Automated systems comprise business rules that cannot be modified without human intervention. This is distinct from *autonomous* systems, which have pre-determined goals but no pre-determined approaches to how they are achieved.

Data analytics (including advanced analytics): The pursuit of extracting meaning from large volumes of (sometimes high-velocity) data using specialized computer systems. These systems transform, organize, and model the data to draw conclusions and identify patterns.

Data governance: Efforts at a corporate level to formally manage data assets (e.g. who does what, when, and under what authority). Effective data governance ensures that data is of high-quality as it is created, extracted and manipulated for reporting and decision-making. A bridge between the business and IT spheres, data governance supports process innovation and improved client service.

"Fettering" of a decision-maker's discretion: Fettering occurs when a decision-maker does not genuinely exercise independent judgment in a matter. This can occur when a decision-maker binds him/herself to a fixed rule of policy, another person's opinion, or the outputs of a decision support system. Although an administrative decision-maker may properly be influenced by policy considerations and other factors, he or she must put his or her mind to the specific circumstances of the case and not focus blindly on one input (e.g. a risk score provided by an algorithmic system) to the exclusion of other relevant factors.

Procedural fairness: Crucial to administrative decision-making, procedural fairness requires that applicants:
— have meaningful participation (they know the case to be met and have an opportunity to meet it) and are provided with a fair and unbiased assessment of their application
— are informed of negative evidence and the decision-maker's concerns
— have a meaningful opportunity to provide a response to such evidence and concerns about their application
— are provided reasons for the decision

There are numerous factors that determine what, specifically, should be done to address these principles. One such factor is the importance of the decision from the applicant's perspective.

Process automation: Also called "business automation" (and sometimes even "digital transformation"), process automation is the use of digital technology to perform routine business processes in a workflow. Process automation can streamline a business for simplicity and improve productivity by taking mundane repetitive tasks from humans and giving them to machines that can do them faster. A wide variety of activities can be automated, or more often, partially automated, with human intervention maintained at strategic points within workflows. In the domain of administrative decision-making at IRCC, "process automation" is used in contrast with "automated decision support," the former referring to straightforward administrative tasks and the latter reserved for activities involving some degree of judgment.

Machine learning: A sub-category of artificial intelligence, machine learning refers to algorithms and statistical models that learn and improve from examples, data, and experience, rather than following pre-programmed rules. Machine learning systems effectively perform a specific task without using explicit instructions, relying on models and inference instead.

Source code: Computer program in its original programming language, human readable, before translation into object code. It consists of algorithms and computer instructions, and may include developer's comments.

25